Sensing cell states induced by chemical probes and approved drugs using fluorescent based biological reporters

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt beim Fachbereich 14

der Johann Wolfgang-Goethe-Universität

in Frankfurt am Main

von

Amelie Menge

aus Frankfurt am Main

Frankfurt am Main 2023

(D30)

"A picture is worth a thousand words"

– Arthur Brisbane

"The dose makes the poison"

– Paracelsus

Vom Fachbereich 14 der

Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Clemens Glaubitz

Gutachter: Prof. Dr. Stefan Knapp, Prof. Dr. Dr. Achim Schmidtko

Datum der Disputation: 07.05.2024

Acknowledgements

I would like to thank everyone who provided both professional and personal support, contributing significantly to the successful completion of this PhD thesis.

First, I want to express my sincere gratitude to Prof. Dr. Stefan Knapp for the opportunity to work in his exceptional research group, filled with open-hearted, science-loving, motivated and talented individuals, making it a truly inspiring and enriching experience. Thank you for your patient, helpful and kind advice, which has helped me grow as a researcher.

Second, I would like to thank my group leader Dr. Susanne Müller-Knapp, without whom this thesis would not have been possible. You taught me the essence of critical thinking, gave me the right push towards my goals, and encouraged and helped me turn my own ideas into successful scientific projects. I would also like to thank Martin Schröder, who equipped me with a lot of helpful lab techniques and has always been by my side with advice, support and good ideas. Additionally, I want to thank Prof. Dr. Dr. Achim Schmidtko for taking the time to evaluate this thesis.

Throughout this experience, I had the privilege of working with many fantastic people. I want to especially thank, Dr. Benedict-Tilman Berger for being a great team leader, for always making me feel valued and for being someone I could turn to with any questions. Special thanks to Dr. Lena Berger. I couldn't have asked for a better office partner. Weather in the lab or in the office, I could always rely on your opinion and support. I wouldn't have made it through this time, without you. The same goes for my colleagues Martin Schwalm and Lewis Elson. Thank you, Martin for showing me every day how fascinating science can be and for inspiring me with your passion and hard work. Thank you Lewis, for being someone that I could count on for support, for never expecting anything in return and for motivating me to learn new languages. Thanks to Joshua Gerninghaus, Theresa Ehret, Julia Frischkorn, Viktoria Morasch, Nikos To, Krishna Saxena, Frederic Farges, Christopher Lenz, Christian Munoz Sosa, Andreas Krämer, Theresa Mensing, Nicolai Raig, Marcel Rak, Nicolas Bauer and the whole Knapp team for great scientific collaborations, interesting journal clubs, energizing lunches and fantastic times during social events including bowling, eating ice cream or Christmas parties.

I also want to acknowledge my colleagues from the SGC Frankfurt, the Institute of pharmaceutical Chemistry, the members of the EUbOPEN project pillar 1 and the people from the third floor of the Buchmann Institut for Molecular Life Sciences (BMLS). Additionally, I want to thank the DFG and the members of the SFB1177 for my funding over the years and the nice experience on retreats and IRTG sessions.

I want to express my gratitude to all collaboration partners with whom I worked during this time, providing me with the opportunity to delve into various scientific fields. Especially, Huabin Hu and Albert Antolin for showing me how fascinating data science can be and Robert Giessmann for making my life much easier by introducing me into the world of programming.

A heartfelt thanks goes to Annika Balzulat, Lena Mareczek, Laura Eichenlaub and Jennifer Dikel, who always listened, helped me maintain a clear mind and encouraged me to approach my challenges from a different perspective.

Finally, I would like to thank the most important people in my life. Thanks to my parents Marco Tjaden and Beate Tjaden, to my brother Lukas and his wife Carina and my grandmother, for their constant support, helpful tips, generous supply of tasty meals and their love. Special thanks goes to my husband, Benni Menge, for his unconditional love and support, for finding the right words after every working day and for giving me the feeling that I can achieve anything I aim for, always having his support.

Table of Contents

Table of Contents

1. Abstract

Bioactive small molecules are used in many research areas as important tools to uncover biological pathways, interpret phenotypic changes, deconvolute protein functions and explore new therapeutic strategies in disease relevant cellular model systems. To unlock the full potential of these small molecules and to ensure reliability of results obtained in cellular assays, it is crucial to understand the properties of these small molecules. These properties encompass their activity and potency on their designated target(s), their selectivity towards unintended off-targets and their phenotypic effects in a cellular system. Approved drugs often engage with multiple targets, which can be beneficial for some applications such as treatment of cancer where several pathways need to be inhibited for treatment efficacy. However, targeting multiple key proteins in diverse pathways also increases the possibility for unspecific or unwanted side effects. For many drugs the entire target space that they modulate is not known. This makes it difficult to use these drugs for target deconvolution or functional assays with the aim to understand the underlying biological processes. In contrast to drugs, for mechanistic studies, a good alternative are chemical tool compounds so called chemical probes that are usually exclusively selective as well as chemogenomic compounds, that inhibit several targets but have narrow selectivity profiles. Because they are mechanistic tools, chemical tool compounds must meet stringent quality criteria and they are therefore well characterized in terms of their potency, selectivity and cellular on-target activity. To ensure that an observed phenotypic effect caused by a compound can be attributed to the described target(s), it is essential to study also properties of chemical tools leading to unspecific cellular effects. There are a variety of unspecific effects that can be caused by physiochemical compound properties that can interfere with phenotypic assays as well as functional compound evaluations. One of these effects is low solubility causing toxicity or intrinsic fluorescence potentially interfering with assay readouts. But unanticipated cellular responses can also arise from unspecific binding, accumulation in cellular compartments or damage caused to organelles such as mitochondria or the cytoskeleton that can result in the induction of diverse forms of cell death.

Abstract

In this study, we investigated the influence of a variety of small molecules on distinct cell states, by establishing and validating high-content imaging assays, which we called *Multiplex* assay. This assay portfolio enabled us to detect different cellular responses using diverse fluorescent reporters, such as the influence of a compound on cell viability, induction of cell death programs and modulation of the cell cycle. Additionally, general compound properties such as precipitation and intrinsic fluorescence were simultaneously detected. The assay is adaptable to assess other cellular properties of interest, such as mitochondrial health, changes in cytoskeletal morphology or phospholipidosis. A significant advantage of the assay is that we are using live cells, so we can capture dynamic cellular changes and fluctuations that can be crucial for the understanding of cellular responses.

Further, I present in this thesis, how we used the assay to annotate different compounds that have been developed in diverse projects of our research team as well as our general effort characterizing chemical tool collections. For instance, in order to evaluate the impact on cellular viability and phenotypic changes of chemogenomic compounds identified by the EUbOPEN project, we performed a two-step viability assessment, consisting of a primary viability assessment based on brightfield microscopy, followed by the high-content *Multiplex* assay for compounds that induced significant cell death or growth retardation. With the established work flow, we contributed to the generation and annotation of the larges available set of high-quality chemical modulators targeting the human proteome, by evaluating the properties of over 1000 diverse chemical compounds. We also enhanced data availability by providing all generated data through open science repositories and helped with the reproducibility of the data by publishing detailed protocols and two step-bye-step method based book chapters.

In one project we examined 150 chemical probes and controls provided by the donated chemical probe (DCP) program assessing the cell viability and phenotypic effects across different standard cancer cell lines. The data generated in 2D tissue culture can be now compared with phenotypic data measured in more complex organoid models. In two probe development projects, we annotated the macrocyclic scaffold compound CK156 and the compound SR-302 with respect of their influence on cell viability and phenotypic changes, which contributed to their approval as

chemical probes for the Serine/threonine kinase 17A (STK17A) (SK156) and for SR-302 as a dual Discoidin domain receptor (DDR)/p38 chemical probe. In another project, we evaluated the effect of Cyclin-dependent kinase 16 (CDK16) on the cell cycle, testing different compounds targeting the PCTAIRE kinase family.

Lastly, one of my main projects was to establish high content assays detecting the poorly understood phenotypic response that causes accumulation of lipids in the endoplasmic reticulum (ER) and lysosomes called phospholipidosis (PL). After the assay was validated, I used a test set of 290 small molecules to generate a high quality data set for the development of a machine-learning based algorithm to identify structural features, critical for a PL induction and can be used further to validate various chemical tool compounds.

In conclusion, my work was mainly focused on the development of high content imaging assays that can now be used for the annotation of chemical tool compounds in a diverse arrange of assays. The established assay portfolio enabled us to improve quality, availability and reproducibility of small molecule data, which is a small, yet significant step towards improving the drug development process and understanding of the underlying biological functions.

2. Zusammenfassung (German Summary)

Die Verwendung bioaktiver Verbindungen, in der Regel kleinmolekularer Substanzen, ist ein wesentlicher Bestandteil der menschlichen Kultur. Diese Verbindungen stellen wichtige Werkzeuge bei der Behandlung von Krankheiten dar und spielen gleichzeitig eine bedeutende Rolle bei der Entschlüsselung biologischer Hintergründe verschiedener Gesundheitszustände[1]. Krankheiten können diverse Ursprünge haben, darunter Infektionen, genetische Faktoren oder Funktionsstörungen des Immunsystems[2]. Seit den frühesten menschlichen Aufzeichnungen wurden Substanzen, die als Bestandteile in Pflanzen oder Materialien aus der Umgebung gefunden wurden, genutzt um Krankheitssymptome zu behandeln. Oft führte dieses Ausprobieren jedoch zu erheblichen Nebenwirkungen, in vielen Fällen sogar zum Tod[3]. Seit dem 19. Jahrhundert hat die *„Pharmazie"* als Disziplin, was die Untersuchung und Erforschung von Substanzen für die Behandlung verschiedener Krankheiten an geht, weiter an Bedeutung gewonnen. Maßgeblich dazu beigetragen hat auch, dass in dieser Zeit bedeutende Fortschritte in der organischen Synthese und in der Isolierung von Verbindungen zum Beispiel aus Pflanzen, gemacht wurden, was das Wissen über molekulare Strukturen von bioaktiven Verbindungen vergrößerte und den Einsatz der Substanzen als Werkzeuge zur Erforschung wichtiger biologischer Funktionen weiter voran trieb[1]. Darüber hinaus haben die Zugänglichkeiten von aktiven Verbindungen und die Möglichkeit, ihre Struktur zu modifizieren, erheblich zur Entwicklung der heutigen Pharmakotherapie beigetragen[1].

Neben der Entwicklung von Arzneistoffen, werden bioaktive Kleinmoleküle in verschiedenen Forschungsbereichen unter anderem eingesetzt um biologische Signalwege aufzudecken, phänotypische Veränderungen zu interpretieren oder Proteinfunktionen zu entschlüsseln. Um hierbei die korrekte Interpretation der resultierenden Ergebnisse zu gewährleisten, ist es entscheidend, eine genaue Kenntniss über die Eigenschaften der Verbindungen zu haben. Dazu gehören das Wissen über die biologische Aktivität der Verbindungen gegenüber ihrem vorher definierten Zielprotein („*Target*"), die Selektivität gegenüber anderen, ungewollten Zielproteinen und ihre phänotypischen Effekte in biologischen Systemen, wie zum Beispiel in Zellen[4]. Die meisten bereits zugelassenen Arzneistoffe interagieren oft mit mehreren Zielproteinen, was vorteilhaft sein kann in Indikationsgebieten wie der Onkologie, wo erst die Inhibierung diverser Signalwege zum erhofften Ergebnis führt. Die Ausrichtung auf

verschiedene Zielproteine vergrößert jedoch auch die Möglichkeit für unerwünschte Nebenwirkungen oder unspezifische Effekte der Kleinmoleküle.

Für die meisten Arzneistoffe sind die Zielproteine, die von den Substanzen moduliert werden, nicht einmal bekannt. Das erschwert die Nutzung dieser Verbindungen, um neue Krankheitsbilder zu therapieren oder funktionelle experimentelle Versuche, zum tiefergehenden Verständnis der zugrundeliegenden biologischen Prozesse, durchzuführen. Eine gute Alternative zur Durchführung solcher mechanistischen Studien ist die Verwendung von so genannten „*chemical tool compounds*"[5]. Diese chemischen Modulatoren sind zellaktive Kleinmoleküle, die eingeteilt werden können in „*chemical probes*", mit besonders hoher Selektivität oder „*chemogenomic compounds*", die mehrere Proteine gleichzeitig modulieren können[6-8]. Damit solche Kleinmoleküle als mechanistische Werkzeuge eingesetzt werden können, müssen sie bestimmte Kriterien erfüllen und daher basierend auf ihrer Potenz, Selektivität und Aktivität in Zellen gut charakterisiert werden[9, 10]. Um sichzustellen, dass die nach Behandlung beobachteten phänotypischen Effekte, die durch diese Verbindungen in späteren experimentellen Versuchen ausgelöst wurden, Rückschlüsse auf das Zielprotein erlauben, ist es wichtig unspezifische Effekte der Verbindungen, auf die in den Experimenten verwendeten Zellen, vorher zu untersuchen.

Es gibt eine Vielzahl unspezifischer Effekte, die nicht nur phänotypische Versuche, sondern auch funktionelle Validierungen der Verbindungen beeinträchtigen können. Unter anderem können physikochemische Eigenschaften der Kleinmoleküle, wie eine geringe Löslichkeit oder intrinsische Fluoreszenz, verschiedene zelluläre Funktionen beeinträchtigen oder die Auswertung der Ergebnisse stören[11-13]. Außerdem können ungewollte zelluläre Reaktionen, die zu einer Induktion diverser Zelltotformen führen können, auch aufgrund unspezifischer Bindungen, einer Akkumulation in zellulären Kompartimenten oder einer Schädigung von Zellorganen, wie Mitochondrien oder dem Cytoskelett, entstehen[14, 15].

In dieser Doktorarbeit, habe ich den Einfluss verschiedener chemischer Verbindungen auf verschiedene Zellzustände untersucht, indem wir ein Hochdurchsatz-Bildgebungsverfahren namens „*Multiplex*" etabliert und validiert haben. Diese Methode ermöglichte es uns verschiedene zelluläre Reaktionen, wie den Einfluss der chemischen Verbindungen auf die Viabilität der Zellen, die Induzierung unterschiedlicher Formen des Zelltodes oder Modulationen des Zellzyklus, zu detektieren, indem wir verschiedene Zelllinien mit

fluorezierenden biologischen Reportern behandelt haben. Gleichzeitig können Eigenschaften der Verbindungen, wie eine Präzipitation in den Zellen oder eine intrinsische Fluoreszenz erfasst werden. Die Methode ist außerdem einfach anzupassen, um andere interessante zelluläre Eigenschaften zu detektieren, wie beispielsweise den Einfluss auf Mitochondrien, Veränderungen in der Morphologie des Cytoskeletts oder die Akkumulation von Phospolipiden.

Ein wesentlicher Vorteil dieser Methode besteht darin, dass lebende Zellen untersucht wurden. Dies ermöglichte die Untersuchung dynamischer zellulärer Veränderungen, die für das Verständnis zellulärer Reaktionen entscheidend sein können. Selbst geringfügige Veränderungen in der Konzentration einer chemischen Verbindung oder in der Expositionszeit, können bereits einen Einfluss auf die zelluläre Antwort haben. Durch die Messung zu verschiedenen Zeitpunkten und bei verschiedenen Konzentrationen ist es möglich zu unterscheiden ob ein phänotypischer Effekt durch unspezifische Zellschädigungen verursacht wurde oder ob Zellschädigungn durch einen phänotypischen Effekt, wie etwa eine Inhibierung des Mikrotubulin-Aufbaus entstanden ist.

Mit Hilfe eines *machine-learning* basierten Algorithmus, teilen wir die Zellen in einer Einzelzellanalyse nach einer Baumstruktur, in verschiedene Gruppen, ein. Mit dieser Methode werden nur Zellen auf phänotypische Eigenschaften untersucht, die in der Analyse vorerst als gesund erkannt wurden. Das trägt dazu bei, die Effekte, die durch Aktivitäts-Eigenschaften der chemischen Verbindung hervorgerufen wurden, von solchen zu unterscheiden, die aufgrund einer unspezifischen Zellschädigung entstanden sind.

Weiterhin beschreibe ich in dieser Arbeit, wie wir das *Multiplex* Testsystem verwendet haben, um verschiedene Verbindungen, die innerhalb unserer Arbeitsgruppe entwickelt oder aufgrund unseres Arbeitsgruppen-weiten Hauptprojekts zusammengestellt wurden, auf ihre Eigenschaften in Zellen hin zu untersuchen. Wir haben den Einfluss 1000 chemischer Modulatoren auf ihre Effekte in verschiedenen Zelllinien untersucht, indem wir eine Zwei-Schritt Evaluierung etabliert und durchgeführt haben. Der erste Schritt umfasst eine mikroskopie-basierte Analyse des Einflusses der Kleinmoleküle auf das Zellwachstum. Alle Kandidaten, die zu einer signifikanten Wachstumsverzögerung geführt haben, welche durch eine Wachstumsrate kleiner 0.5 angezeigt wurde, oder die phänotypisch auffällig waren,

wurden im zweiten Schritt mit der Druchführung des *Multiplex* Testsystems charakterisiert. Mit dieser Vorgehensweise haben wir innerhalb des Arbeitsgruppen-weiten EUbOPEN Projekts, zur Generierung und Beschreibung des größten Sets an Daten für hochwertige chemische Modulatoren des menschlichen Proteoms beigetragen.

Weiterhin haben wir dazu beigetragen, die Verfügbarkeit und Reproduzierbarkeit von Daten für die Wissenschaft im Allgemeinen zu verbessern, indem wir die von uns generierten Daten in einer öffentlich zugänglichen Datenbank bereitgestellt haben, sowie ein detailiertes Protokoll und zwei methodische Schritt-für-Schritt Anleitungen in Fachbuchkapiteln veröffentlicht haben.

In einem weiteren Projekt haben wir 150 chemische Modulatoren des „*Donated chemical probes*"-Programmes mit den dazugehörigen Kontrollen, auf ihren Einfluss auf die Viabilität von Zellen, sowie phänotypische Veränderungen in diversen Krebszelllinien, untersucht[16]. Diese Daten, die in 2D-Zellkulturen aufgenommen wurden, konnten anschließend genutzt werden, um sie mit komplexeren 3D-Modellen, wie Organoid-Studien, zu vergleichen.

In zwei Projekten haben wir den Einfluss von Kandidaten für die Zulassung als „*chemical probe*", auf ihren Einfluss auf die Zellviabilität und phänotypische Veränderungen mit Hilfe des hier entwickelten *Multiplex* Testsystems untersucht. Diese Evaluierung hat dazu geführt, dass die makrozyklische Verbindung CK156 für die Serine/Threonin-Kinase 17 A (STK17A) und die Verbindung SR-302 als dualer Discoidin-Domänen-Rezeptor (DDR) und p38 Binder als „*chemical probe*" zugelassen werden konnte.

Das *Multiplex* Testsystem kann in einfacher Weise angepasst werden, um andere zelluläre Reaktionen auf verschiedene Verbindungen zu untersuchen, die für den zu untersuchenden Forschungsbereich von bedeutenderer Rolle sind. Ich habe daher das *Multiplex* Testsystem unter anderem mit dem „*fluorescent ubiquitin cell cycle indicator*" (FUCCI) verbunden, um Veränderungen im Zellzyklus, hervorgerufen durch chemische Verbindungen, zu erkennen[17-19]. Mit Hilfe dieses Versuchsaufbaus konnte dann der Einfluss der cyclin-anhängigen Kinase-16 (CDK16) auf den Zellzyklus untersucht werden, indem wir verschiedene Inhibitoren der PCTAIRE Kinase-Familie getestet haben.

Weiterhin konnten wir das *Multiplex* Testsystem dahingehend anpassen, dass Verbindungen detektiert werden können, die zu einer Akkumulation von Phospholipiden in Geweben,

vornehmlich in Lysosomen, führen. Dieses Phänomen der sogenannten Phospholipidose, kann zu der Zerstörung normaler lysosomaler Funktionen führen und geht mit morphologischen Veränderungen einher, die potentiell Krankheiten der Leber, Lunge und anderer Organe hervorrufen können[20, 21]. Der Mechanismus hinter Phospholipidose ist sehr komplex und bis heute nicht vollständig verstanden. Da Verbindungen, die zu Phospholipidose führen, oft mit Sicherheitsproblemen und Nebenwirkungen für Patienten verbunden sind, ist es wichtig potenzielle Arzneistoffkandidaten daraufhin zu untersuchen. Nachdem das Testsystem etabliert war, haben wir daher 290 Verbindungen getestet, um ein Datenset zu generieren, welches anschließend genutzt wurde, um einen *machine learning* Algorithmus zu etablieren, der strukturelle Eigenschaften von Verbindungen erkennt, die dazu führen, dass eine Verbindung eher zu einer Phospholipidose-Induktion führt. Sowohl der Algorithmus, als auch der das etablierte Testsystem kann nun genutzt werden, um diverse chemische Verbindungen auf Phospholipidose hin zu untersuchen.

Zusammenfassend konnte ich zeigen, dass eine Charakterisierung von chemischen Verbindungen hinsichtlich ihrerer Eigenschaften im zellulären Kontext, nicht nur für chemische Modulatoren, sondern für alle chemischen Verbindungen, von großer Bedeutung ist, wenn man diese für mechanistische Studien und Signalweg-Analysen verwenden möchte. Das hier vorgestellte Hochdurchsatz-Bildgebungsverfahren, welches wir als *Multiplex* Testsystem etabliert haben, ist eine gute Methode, um verschiedene Zellzustände und zelluläre Reaktionen, induziert durch chemische Verbindungen, auf einfache und schnelle Weise zu erkennen. Es bietet zudem breite Anwendungsmöglichkeiten und lässt sich mühelos anpassen, um projektspezifische oder krankheitsbilderbezogene Fragestellungen gezielt zu beantworten. Die Annotierung der Verbindungen und Bereitstellung der Daten trägt maßgeblich dazu bei die Qualität, Verfügbarkeit und Reproduzierbarkeit von Daten zu verschiedenen chemischen Verbindungen zu verbessern, was ein kleiner aber bedeutender Schritt zur Verbesserung des Arzneimittelprozesses und des Verständnisses der zugrundeligenden molekularen Mechanismen der Medikamentenkandidaten ist.

3. Introduction

3.1. Biological targets in drug development

Bioactive small molecules are a fundamental components of human culture, as they are our most important tools for the treatment of human diseases, and also play a significant role in deciphering biological pathways underlying various health conditions[1]. Diseases can occur through a variety of mechanisms, including infections, genetic factors or immune system dysfunctions[2]. Since the earliest record, substances which derived from plants or materials found in the surroundings of early humans, were used to treat disease symptoms. These trial-and-error attempts often resulted in significant side effects or in many cases even led to death[3]. Since the 19th century, the discipline of pharmacy emerged, encompassing the examination and study of substances intended for the treatment of various diseases. At this time, significant advancements were made in organic synthesis and in the field of isolation of compounds, including the capability to extract active substances from natural products such as plants[1]. The knowledge about chemical structures significantly promoted the use of small molecules as research tools for understanding biological functions and dysfunctions. Additionally, the accessibility of active substances and the ability to modify their structure have greatly contributed to the development of the pharmacotherapy as we know it today[1]. The identification and optimization of active substances for the development of a new drug to treat diseases, without harming the organism as a whole, is one of the most crucial scientific challenge facing humanity. The optimization of a small molecule to a drug takes on average 15 years from start to the final clinical phase and costs around 2 billion US dollars[4, 22]. Even after this considerable investment of time and resources, the final approval of the drug candidate as a drug remains uncertain, because the drug discovery process is associated with a high risk of attrition[23]. Out of 20.000 drug candidates entering the lead optimization phase of the drug discovery process, only one will ultimately be successfully being approved for the treatment of a specific disease[4, 24]. To optimize and enhance the entire drug development process, it is of utmost importance to understand the biological origin of the disease one intends to treat. Specific molecules or biological entities (such as proteins, receptors or genes), that are associated with a particular disease or pathological condition are commonly referred to as "*targets*"[4]. Frequently, drug molecules are designed to have a primary target being responsible

for the therapeutic effect. However, most times, drug molecules also influence diverse other proteins or processes, which is a common phenomenon in the field of pharmacology, referred to as "*polypharmacology*". This potentially leads to additional therapeutic benefits or, in some cases, unwanted side effects. Understanding a drug's target interaction is therefore crucial to minimize unanticipated side effects, rationally optimizing the compound and to uncover new uses for existing medications. If the disease-specific biological target is already known, a chemical modulator can be identified using screening efforts or developed by structure-activity relationship approaches. In the drug development process, this target-oriented approach is referred to as the "*reverse approach*" (Figure 1). The opposing approach, called "*forward approach*", starts with a compound, which has been identified in a disease relevant phenotypic screen[25]. The compound is then tested in various systems such as *in vitro* or *in vivo* analyses to identify the target(s) modulated by it. Such compounds are usually found by screening large libraries of compounds[5, 7, 13, 26].
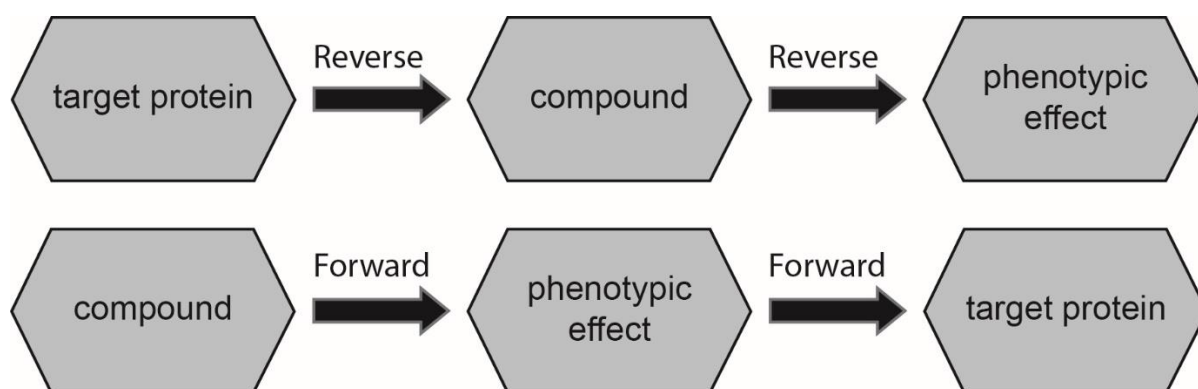
Figure 1: Drug-target identification approaches. Schematic overview of "*forward*" and "*reverse*" approach in the drug development process[1]. This figure was created using Adobe Illustrator v26.5.

3.1.1. Target identification via phenotypic screening

In the forward approach, phenotypic screening has regained interest in the field of drug development in recent years[25, 27]. Rather than focusing on specific molecular targets or substrates, this approach evaluates desired observable changes or phenotypes caused by a compound in a biological system. The number of drugs approved by the Food and Drug Administration (FDA) in recent decades was higher for drugs discovered by phenotypic

screening compared to target-based approaches[28]. If disease-relevant systems are available for compound screening, phenotypic screening is often the method of choice[7]. A critical aspect of phenotypic screening is the choice of the disease-relevant cellular model system used. Over the past decades, comprehensive collections of cellular model systems were curated by so called "*cellbanks*" with thousands of choices for immortalized cell lines relevant in different diseases or disease states[29]. Each cell line exhibits a unique protein expression profile and signaling pathways based on tissue types, different diseases or relevant mutations, resulting in a variety of phenotypic responses when exposed to different small molecules. To obtain a comprehensive understanding of the potential interactions of a compound within the cellular system, it is essential to test them in multiple cell lines representing different tissues. For more disease-related inquiries, using specific patient-derived cells, that are more likely to be clinically relevant, can further enhance the accuracy and applicability of the screening approach[30, 31].

Furthermore, compounds that were identified using phenotypic screening technologies, often demonstrate the effect of the most phenotypically relevant target in the pathway which means the protein they bind with the highest potency. They are also more likely to act on multiple targets (polypharmacology), leading to off-target effects, which can result in diverse adverse events[32]. Therefore, a retrospective identification of the responsible target or a deconvolution of multiple targets is often limited[1].

A simpler approach to identify targets responsible for a physiological response is the use of highly selective and well validated tool compounds called chemical probes[33, 34] or a recently emerged strategy called chemogenomics[35].

3.1.2. Target deconvolution using chemical tool compounds

Approved drugs often have multiple targets, which can be beneficial when target modulation results in a broader spectrum of efficacies, either affecting multiple proteins important for the same disease or allowing a drug to address other diseases as well. On the other hand, these different protein-target interactions often lead to safety problems of the drug. In most cases, the primary target leading to the desired pharmacological effect is not known (Figure 2 A). Many clinical trials have failed and wrong conclusions about biological processes have been drawn from approved drugs that lack characterization data[33]. For example the v-raf murine sarcoma viral oncogene homolog B1 (BRAF) kinase inhibitor Dabrafenib, approved by the FDA

for the treatment of unresectable or metastatic BRAF-mutated melanoma, advanced non-small cell lung cancer and anaplastic thyroid cancer[36, 37], was shown to also have activity on other kinases such as ALK5 (Activin receptor-like kinase 5), LIMK1 (LIM domain kinase 1), NEK11(NIMA-related kinase 11) and SIK1(Salt-inducible kinase 1) enhancing the probability for adverse events[37, 38]. Another example is Crizotinib, originally developed as cMET (mesenchymal-epithelial transition factor) inhibitor, which has been approved by the FDA and EMA for the treatment of metastatic non-small cell lung cancer (NSCLC) with anaplastic lymphoma kinase (ALK) rearrangements, as this inhibitor was shown to also target ALK and ROS proto-oncogene1 (ROS1) that are deregulated in a subtype of NSCLC[39-41].

In contrast to drugs that inhibit multiple targets, chemical tool compounds offer a versatile strategy to understand the specific target functions at the mechanistic level[5]. Chemical probes are cell-active small molecules that selectively modulate the function of a specific protein[8] (Figure 2 B). The modulation of a protein function for an enzyme would be an inhibition of its biochemical reaction. The affinity of a potential inhibitor is a measure of the strength of this inhibition effect and is measured in the dissociation constant. The dissociation constant of the enzyme-inhibitor system is defined as the concentration of inhibitor that inhibits the enzymes function to 50 % of its maximal effect[42]. Hence, the lower the dissociation constant, the less inhibitor is needed to inhibit the enzyme or the more potent the inhibitor is. Chemical probes are highly validated and should meet stringent criteria for potency, selectivity and toxicity. Precisely, they need an *in vitro* potency lower than 100 nM in a biochemical or biophysical assay or lower than 1 µM in cells, a 30-fold higher activity over related proteins and must not be cytotoxic at 10 µM[8, 10, 43]. When well profiled, chemical probes can be used for mechanistic and phenotypic studies to decipher the biology of their target at the appropriate concentration[5, 10, 44]. Complementary to genetic tools such as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and ribonucleic acid (RNA) interference, they can rapidly alter a specific protein's function and improve the knowledge about the connection between genomic target identification and druggability of a target[7, 45, 46].

When using chemical probes, it is good practice to use them at the recommended concentrations and to use a structural closely related inactive control, to confirm that the observed effect is caused by the main target and not due to unspecific binding or off-target effects beyond the 30-fold selectivity window. The non-existence of certain characteristics

such as selectivity or unknown binding to cellular structures may cause misleading results and result in conflicting data about the function of a target and its role in biology[43]. It is therefore important that compounds are profiled comprehensively and that this data is made available[47]. In the last decade, academic laboratories in close collaboration with pharmaceutical companies have joined forces and have helped to systematically enlarge the number of available chemical probes e.g. for the target family of protein kinases[7, 34]. Initiatives such as the Chemical Probes Portal (www.chemicalprobes.org)[10], an online resource for sharing data and reviewing various compounds[9, 48] and the chemical probes website of the Structural Genomics Consortium (SGC) (https://www.thesgc.org/chemical-probes), a public-private consortium of academical institutes and pharmaceutical companies that offer access to diverse sets of chemical probes as well as the Donated Chemical Probes (DCP) program (https://www.sgc-ffm.uni-frankfurt.de/)[43], have helped to significantly improve the quality of tool compounds and their availability to the global research community[43, 49]. As it is often challenging to develop inhibitors targeting one specific protein, a new approach, called chemogenomics, has gained interest in recent years.

In contrast to chemical probes, chemogenomic compounds target often more than one protein and preferably represent different chemotypes[6, 49, 50]. The compounds included in chemogenomic libraries should have known selectivity profiles and distinct mode of actions [51]. Compounds with overlapping target spectra are tested in parallel. Based on the observed phenotypic effect and these tight target spectra, deconvolution conclusions can then be drawn about the responsible target (Figure 2 C). For building such a library and being able to connect a specific chemotype to a target it is crucial to evaluate a compound on diverse criteria like selectivity, potency as well as their effect in a cellular context[44]. Using diverse chemogenomic libraries thus enlarges the targetable space, which is needed due to the unavailability of chemical probes for every target.
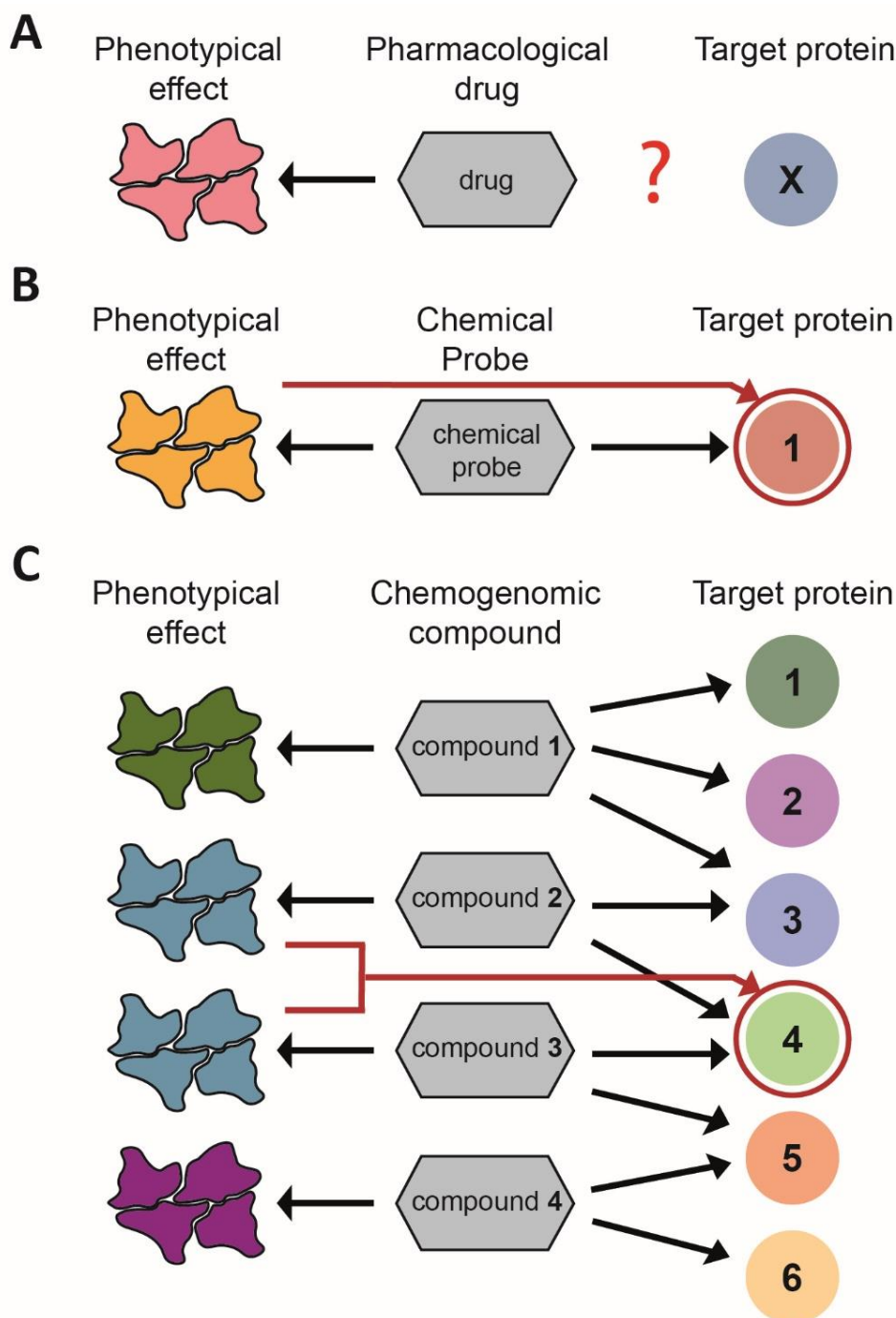
Figure 2: Graphical Illustration of target-effect relationship of pharmacological drugs, chemical probes and chemogenomic compounds. **A.** Approved drugs are frequently identified through phenotypic screening. These drugs often interact with multiple proteins and their target might be unknown. **B.** Chemical probes are highly validated, cell active small molecules that selectively target one specific protein. The resulting phenotypical effect can be traced back to the main target protein. **C.** Chemogenomic compounds have more than one target. They are used within a set of well-validated and characterized compounds. By examining the observed phenotypical effect of diverse compounds, conclusions can be drawn about the responsible target for the mode of action. This figure was created using Adobe Illustrator v26.5.

3.2. Quality and Properties of small molecules in cellular systems

Small molecule modulators play a crucial role in drug discovery and target identification. They are more versatile than genetic tools as they can be applied in a dose-and–time dependent manner, and be used in cell models that are more difficult to address with genetic tools[49]. Moreover, specific protein domains can be targeted enabling modulation of a protein's function without removing the protein and its interactome. However, small molecule modulators need to be well-characterized to allow linking the phenotype to a specific target or group of targets. Unfortunately, in the past decades, a large number of chemical compounds were published that were not properly validated. Often, the global research community retrospectively reported these chemical compounds to be non-selective, to contain chemical or physical properties interfering with different assays[5] or to have off-target activities leading to cytotoxic and other effects on various cellular compartments that could not be linked to the intended target[44, 52].

In 2010, Baell *et al.*[11] highlighted inherent flaws of certain compounds, which interfered with biological assays, making it challenging to accurately assess the true activity of the compounds. These Pan-Assay Interference Compounds (PAINS) often exhibit different effects such as toxicity, reduced efficacy and unwanted side effects through off-target or unspecific binding[53]. For example, these compounds were often characterized by their propensity to exhibit promiscuous binding behavior to various targets[11]. Also chemical substances could show biological effects through diverse mechanisms, such as non-specific aggregation, redox activity or fluorescence interference. Certain compounds are also classified as PAINS based on the structural properties of the compound itself[12, 52]. It is therefore crucial to validate compounds at both, the chemical and cellular effect level and make this data available to the public. Although computational PAINS filter are available to help the researchers, ultimately, PAINS have to be determined for each assay used[12].

3.2.1. Chemical quality of a compound can influence cell reactions

An underappreciated problem is that commercially available compounds are not always what they are supposed to be. This ranges from different salt forms to different compound identities such as different isomers, which may be inactive on the supposed target or a complete mix of different impurities or heterogeneities. Moreover, the purity of the compounds may not be

sufficiently high and influence a cellular readout as the impurities may influence cell viability or interfere with the assay. Therefore, ideally the identity of compounds should be confirmed and verified through appropriate analytical techniques such as spectroscopy, chromatography, or mass spectrometry[54, 55]. Also chemical and metabolic stability under common storage and handling conditions should be guaranteed[56, 57]. The compounds shouldn't undergo degradation, decomposition, or reaction with environmental factors such as water or oxygen under standard light or temperature conditions. One significant aspect of chemical stability also includes the solubility of a compound, which refers to the compound's capacity to dissolve in a specific solvent or solvent system[58]. Additionally, solubility can influence the safety and toxicity profile of compounds, as poorly soluble ones may tend to accumulate within cells or tissues.

3.2.2. Methods to assess cell viability

One of the prevalent effects exhibited by compounds is their impact on viability. It is crucial to differentiate between the desired effect, based on the compound's on-target activity and its off-target activity. For instance, in cancer treatment, the standard therapy relies on selectively killing dangerous, uncontrollably growing cells while preserving normal cells[59]. Viability refers to the overall health and functionality of cells or organisms. It indicates whether cells are alive and capable of normal physiological processes[60]. Therefore, standard viability assays assess properties such as metabolic activity and other indicators of cellular activity.

Assessing the impact of a compound on cell viability or proliferation can be conveniently and rapidly done using a variety of end-point and commercially available viability assays. These assays operate through diverse mechanisms, so it is advantageous to perform a combination of them. One commonly employed mechanism is the assessment of the metabolic activity of cells using different reagents that can be detected via a spectrometer or luminescence reader. A frequently used assay involves the utilization of tetrazolium salts, such as MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay[61] or XTT (2,3-bis[2-methoxy-4-nitro-5-sulfophenyl]-2H-tetrazolium-5-carboxanilide)[62-65]. In this approach, the activity of mitochondrial dehydrogenase in metabolic active cells is assessed, which converts the yellow tetrazolium salt to a visually detectable purple formazan product (Figure 3 A)[66]. The amount of formazan product can be quantified using a spectrophotometer, and correlates with the viability of the cells[63]. A similar information regarding proliferation, with less incompatibility

problems such as color interferences of the used compounds or toxicity of the salt[67] can be obtained using the more sensitive alternative AlamarBlue™, also known as resazurin[68-70] (Figure 3 B). Metabolically active cells are able to reduce the non-fluorescent resazurin to a highly-fluorescent red product called resorufin, which can be detected spectrophotometrically and allows conclusions regarding the metabolic activity of the cells[67]. Another mechanism is based on the measurement of cellular adenosine triphosphate (ATP), which is present in metabolically active cells and a widely accepted marker of viable cells[71]. These assays for example the CellTiter-Glo® assay (Promega) take advantage of the firefly luciferase enzymatic reaction, which requires ATP present in viable cells to generate a luminescence signal (Figure 3 C)[72, 73]. The intensity of the luminescent signal can be detected using a luminometer, with higher luminescent signal indicating higher ATP levels and higher viability of the cells[71].

Viability assays based on spectrometers or luminescence read-out provide a rough estimate of the impact of compounds on cell proliferation. Nevertheless, they do not encompass the full complexity of the underlying cellular mechanisms or distinguish between viability and cytotoxicity[74].
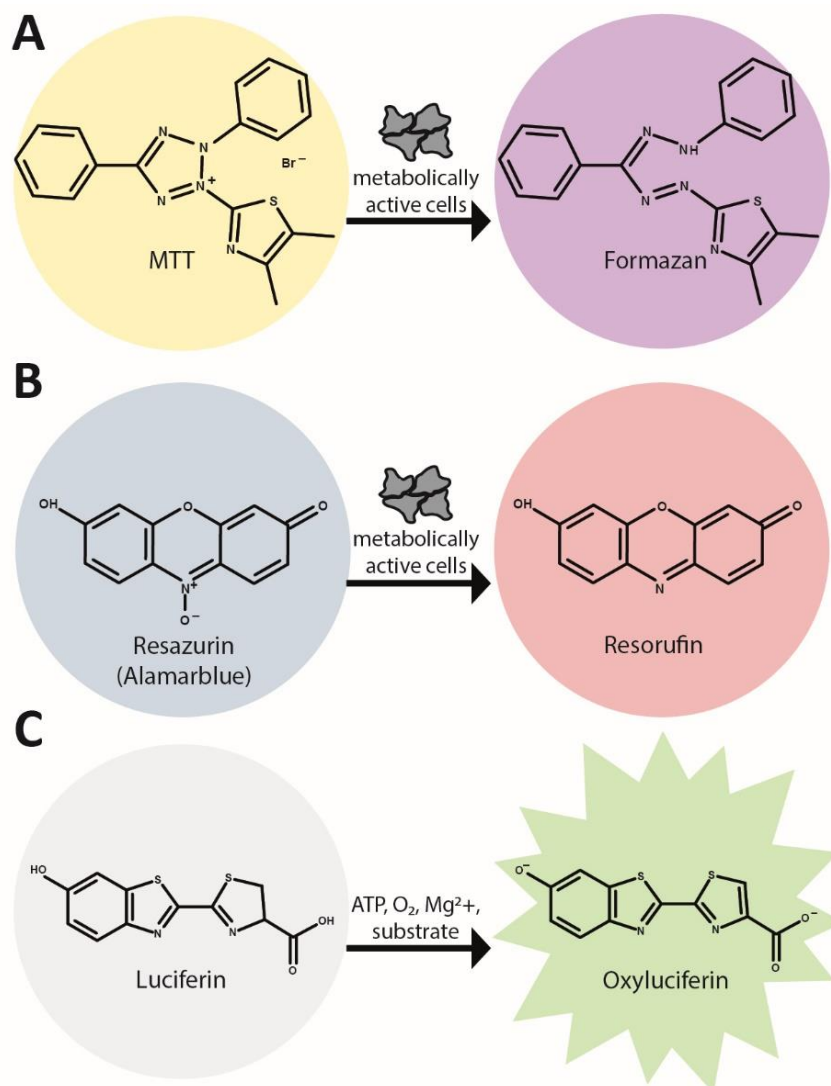
Figure 3: Graphical illustration of end-point viability and cytotoxicity assessment. **A.** MTT assay. The yellow tetrazolium salt MTT is reduced by metabolic active cells to formazan (purple). **B.** AlamarBlue™ assay. The non-fluorescent resazurin is reduced by metabolic active cells to red fluorescent resorufin. **C.** CellTiter-Glo® assay. Based on ATP, which is present in living cells, Luciferin can be oxidized to oxyluciferin which generates a luminescence signal. This figure was created using Adobe Illustrator v26.5.

3.2.3. Target associated and non-specific cytotoxicity

Specific cytotoxicity refers to the selective destruction of specific cells or tissues while sparing others. This approach is commonly employed in cancer therapy, where the goal is to eliminate cancer cells without harming healthy, normal cells[75]. Target-specific cytotoxicity is achieved by using therapies that specifically target molecular markers or vulnerabilities unique to the cancer cells. Unspecific cytotoxicity, on the other hand, refers to the non-selective destruction of cellular functions without a specific focus on any cell type where unintended damage to

both diseased and healthy cells can occur[76, 77]. This type of cytotoxicity is often associated with the side effects of certain drugs or therapies, particularly chemotherapy and is caused by toxic effects of compounds[78]. Cell death is a highly complex and dynamic process involving several different mechanisms. Each cell death mechanism has distinct molecular and morphological features and serves different fundamental biological processes. Understanding the influence of a compound on each mechanism is crucial for unraveling the compounds' profile and gaining further target related insights. The best studied forms of cell death include apoptosis, the programmed cell death, and necrosis, the uncontrolled cell death form. However, in recent years several other forms of cell death have been discovered for example autophagy, necroptosis and oncosis[77].

3.2.3.1. Programmed cell death: Apoptosis

Apoptosis, also known as programmed cell death, is a highly regulated and controlled process to eliminate unwanted or damaged cells to regulate cellular growth, differentiation and homeostasis (Figure 4 A). It is characterized by distinct morphological changes of the cells, starting with shrinkage of the cells and condensation of the cytoplasm combined with nuclear fragmentation and formation of apoptotic bodies[79, 80]. Apoptotic cells are subsequently phagocytosed or cleared by neighboring cells without inducing inflammation[80, 81]. To initiate apoptosis, frequently a series of proteases known as caspases is activated. Caspases are aspartate-directed cysteine proteases that are able to cleave diverse intracellular substrates. They are synthesized as inactive precursors and can be grouped into initiator caspases (caspase 8 and caspase 9), that transduce various signals into proteolytic activity and executioner caspases (caspase 3, caspase 6 and caspase 7), which cleave most of the more than 400 substrates that are degraded in cells undergoing apoptosis[77, 82, 83]. They are activated as a consequence of signaling induced by a wide range of physiological and pathological stimuli and contribute to various manifestations of apoptosis[83]. Apoptosis can be initiated by two different pathways: the intrinsic pathway and the extrinsic pathway. The intrinsic pathway is characterized by mitochondrial membrane changes due to positive or negative signals[77]. Negative signals result from the lack of various cytokines, hormones, or growth factors. Within non-apoptotic cells, these factors inhibit the activation of pro-survival or pro-apoptotic factors such as puma, noxa, or bax[84]. Once triggered, these elements have the potential to initiate apoptosis. The positive signals include external influences such as hypoxia, different toxins,

radiation, viruses or toxic agents[77]. Upon exposure to an apoptotic stimulus, the mitochondrial membrane changes to open a mitochondrial permeability transition (MPT) pore and pro-apoptotic proteins such as cytochrome c, second mitochondria derived activator of caspases/direct IAP binding protein with low PI (Smac/Diablo) or the high-temperature requirement family member HtrA2/Omi can leak into the cytoplasm, where they can activate executioner caspases[84, 85].

The extrinsic pathway or death receptor (DR) pathway is initiated by patrolling natural killer cells or macrophages. So called death ligands bind to death receptors (DR) which are transmembrane proteins and members of the tumor necrosis factor (TNF) superfamily[86]. By means of their death-inducing domain (DED), they can activate pro-caspase 8 to caspase 8, thereby forming a death-inducing signal complex (DISC). This complex includes various other proteins, that contribute to the autocatalytic activation of caspase 8[80, 82]. Additionally, it leads to the subsequent activation of executioner caspases either directly or in conjunction with proteins released from mitochondria, which counteract the inhibitors of apoptosis proteins (IAPs) from inhibiting caspase 8[80]. This non-inflammatory, highly regulated form of cell death is for example important to clear damaged or infected cells in order to maintain surrounding healthy cells. Nevertheless, in some cases, for example for more selective cell depletion or when a broader area of cells should be removed, other cell death mechanisms are preferred.

3.2.3.2. Energy-independent cell death: Necrosis and Necroptosis

In contrast to the programmed cell death form apoptosis, necrosis is an energy independent form of cell death (Figure 4 B). It is an uncontrolled mechanism, induced by external damaging conditions such as physical trauma, toxins, hypoxia, or sudden inflammation[86]. Necrosis is characterized by leakage of cellular compartments and nucleosides into the surrounding environment, membrane rupture, cellular swelling (often referred to as oncosis) and triggers inflammatory processes[77]. After cell damage occurs, various pro-inflammatory proteins such as nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB) are upregulated, triggering the inflammatory response to remove harmful components. A specific form of necrosis is necroptosis, which is a partially regulated process. It is assumed, that necroptosis happens in an apoptosis deficient environment, controlled by the two receptor-interacting proteins (RIPK) RIPK1 and RIPK3[77]. Necroptosis is mediated by death receptors (DR) such as TNFR1 (tumor necrosis factor receptor 1), which recruits the survival complex 1, containing

TNFR1-associated death domain protein (TRADD), polyubiquinated RIPK1 and several E3 ligases. After deubiquination of RIPK1 it recruits and phosphorylates RIPK3, forming the ripoptosome[87]. This complex recruits and activates the mixed lineage kinase domain such as pseudokinase (MLKL) to form the necrosome, leading to membrane permeabilization and cell death[77, 86].

3.2.3.3. Autophagy

Another mechanism to maintain cellular homeostasis and promoting overall cell health is the self-digestion called autophagy (Figure 4 C)[88]. It is a highly conserved cellular process through which cellular compartments or whole cells are degraded by lysosomal enzymes. After digestion, the components can be recycled or even used as a source of energy, which plays an important role during times of nutrient scarcity, which is next to other internal and external signals, one of the main autophagy triggers[89].

The literature distinguishes between three forms of autophagy: macroautophagy, microautophagy and chaperone-mediated autophagy. The most described form is macroautophagy[90]. Here, the cargo is enclosed in a double-membrane vesicle (autophagosome) and later fused with a lysosome to the autophagolysosome, where the cargo is released into the lysosomal lumen, followed by degradation by lysosomal proteases. The resulting breakdown products, such as amino acids, nucleotides, or fatty acids are released back to the cytoplasm, where they can be reused by the cell for energy production or the synthesis of new cellular components[88]. The more specific microautophagy is so far not fully understood. It is characterized by an incorporation of cytoplasmic content into the lysosome by an invagination of its surface[91, 92]. The most selective form of autophagy is called chaperone-mediated autophagy. It is dependent on cytosolic chaperones, which can bind specific target proteins of interest carrying KFERQ-like motifs and transfer them via the lysosomal-associated membrane protein (LAMP) receptor 2A to lysosomes where they are degraded[92]. All forms of autophagy are highly regulated and influenced by different proteins[93].

Autophagy has often been shown to be induced by chemical compounds. Furthermore, it was also shown to play a crucial role in diverse pathologies including neurodegeneration, cancer and infectious diseases, thus representing as an interesting leverage point for therapeutic strategies[93]. In cancer therapy, autophagy was shown to both suppress and promote cancer

progression and metastasis, dependent on the patient and cancer form[93]. While pharmacological inhibition of autophagy has demonstrated the potential to enhance the effectiveness of cancer therapies and reduce drug resistance, it is important to note that certain types of cancer rely on autophagy for their survival. Additionally, there are instances where boosting autophagy serves as a strategy to improve the efficacy of cancer treatments[94]. The first FDA-approved agent capable of inhibiting autophagy was chloroquine, which blocks autophagy by disrupting lysosomal acidification[93, 95]. Originally developed as a treatment for malaria, this drug garnered significant interest, particularly during the coronavirus pandemic of 2019 (COVID-19), when it became a subject of research and was discussed for potential antiviral therapy[96, 97]. The exact mechanism of this drug, however, are complex and not yet fully understood. In addition to compounds that disrupt lysosomal function, over the years a wide array of compounds has been developed to specifically target essential proteins of the autophagy machinery to either inhibit or induce autophagic functions. An example is the immunosuppressant drug Rapamycin, which targets mammalian target of rapamycin (mTOR), an important serine/threonine kinase and key regulator of the autophagy pathway. Inhibition of mTOR by Rapamycin leads to autophagy induction as active mTOR inhibits autophagy through the inhibitory phosphorylation downstream of unc-51 like autophagy activating kinase 1 (ULK1)[98-100]. Later, the drug was also shown to have potential in different other therapeutic areas such as cancer therapy and aging[100]. Effectively targeting autophagy without causing harm to other essential cellular processes remains a significant challenge in this field. Additionally, it is important to acknowledge that while the design of many compounds focuses on specifically targeting autophagy-related proteins, there are also compounds that can induce autophagy or general cytotoxic effects through mechanisms that are not yet fully understood. Therefore, it is desirable to characterize the impact of a compound on autophagy or cell viability in general, during the early stages of drug discovery.
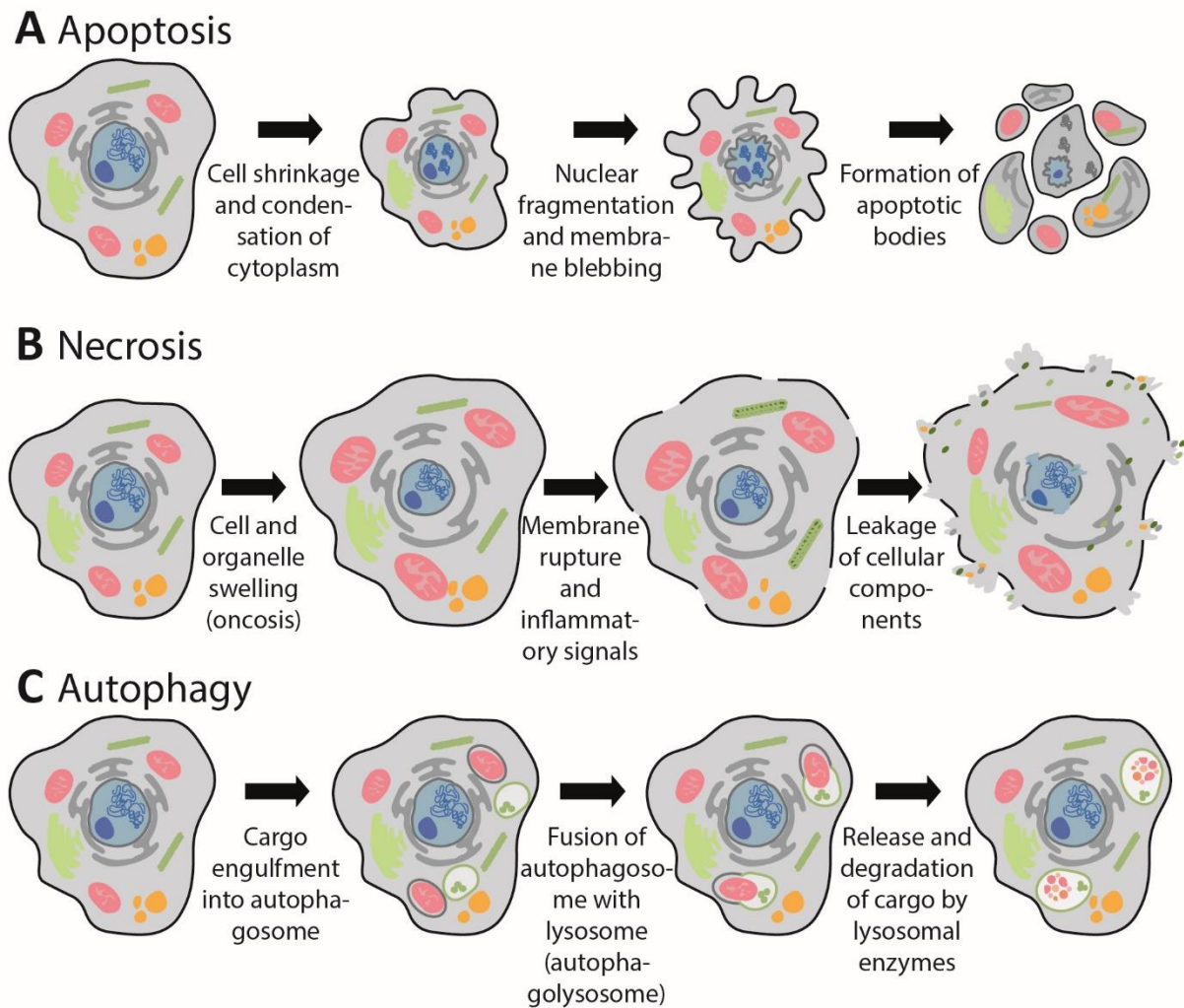
Figure 4: Overview of cell death mechanisms. Figure adapted from Asadzadeh *et al.*[101] **A.** Apoptosis is a highly regulated and controlled process, characterized by distinct morphological changes of the cells, including cell shrinkage and condensation of cytoplasm, nuclear fragmentation and membrane blebbing and the formation of apoptotic bodies. **B.** Necrosis is an energy-independent form of cell death, characterized by cell and organelle swelling (oncosis), membrane rupture, inflammatory processes and leakage of the cellular components into the cytoplasm. It is often induced by external stimuli such as trauma, toxins or sudden inflammation. **C.** Autophagy is a highly conserved process to degrade and recycle cellular components or whole cells. The here shown macroautophagy is characterized by a cargo engulfment in an autophagosome, a fusion of this autophagosome with a lysosome and the release and degradation of the cargo by lysosomal enzymes in the resulting autophagolysosome. This figure was created using Adobe Illustrator v26.5.

3.2.5. Cell cycle dependent effects

Next to inducing different cell death types, some compounds can influence cell viability by impacting the cell cycle. The cell cycle is a highly regulated and precisely controlled process in

cells to ensure accurate progression from one cell cycle phase to another, leading to cell division and proliferation. It consists of four main phases with specific functions and checkpoints, to govern cell growth, development and reproduction of the cell, and maintaining the integrity of the genetic material. The gap 1-phase (G1 phase) is the interval phase between mitosis (M phase) and Deoxyribonucleic acid (DNA) synthesis (S phase)[102] (Figure 5 A). Here, the cell is growing and prepares itself for DNA replication. Cellular proteins, organelles and RNA that are important for the division are synthesized. Regular functions and metabolic activities of the cell continue unaffected[103]. A pivotal moment in this phase involves a decision point at which the cell either enters the cell cycle and proceeds with division or enters the non-dividing state, known as quiescence or G0 phase. This decision is highly regulated and influenced by various factors, such as nutrients, growth factors and the overall cell health. Upon choosing to move forward with the cell cycle progression, the following phase is called DNA synthesis or S phase. Here, the genetic material, which is stored in the form of DNA, is duplicated to form two identical copies of each chromosome. Followed by a second gap phase (G2), where the cell undergoes final preparations for cell division and when a second decision point takes place[102]. It is a critical checkpoint, to ensure that the cell has successfully completed DNA replication and is ready to enter mitosis (M phase). During mitosis, the duplicated DNA is separated and the entire cellular content is divided into two new daughter cells (cytokinesis). Afterwards, the cell cycle returns to G1 phase.

The whole cell cycle progress is controlled by a complex network of regulatory mechanisms and diverse proteins. Key regulators are cyclin-dependent kinases (CDKs)[104, 105]. They interfere and regulate the cell cycle progression by binding to specific cyclins, which are present at different cell cycle phases. Binding to the corresponding cyclin leads to CDK activation and phosphorylation of proteins, triggering specific events of the cell cycle[105, 106].

In many cancer types, cell cycle proteins are frequently overexpressed, leading to uncontrolled proliferation[107]. Inhibition of proteins such as CDKs has demonstrated to block initiation and progression of specific cancer types, without harming healthy tissues. In some instances, cell cycle arrest even triggered apoptotic cell death or senescence of tumor cells[107]. Some compounds exhibit their mode of action effectively within a specific phase of the cell cycle[108]. Overall, this renders proteins engaged in the cell cycle as promising targets for drug

development and highlights the need to assess the impact of compounds on the cell cycle during the drug development process.

There are various techniques to detect the cell cycle such as using nucleotide analogs 5-bromo-2'-deoxyuridine (BrdU) and 5-ethynyl-2'-deoxyuridine (EdU), which are incorporated into newly synthesized DNA, immunostaining of replication proteins such as Ki-67 or proliferating cell nuclear antigen (PCNA), or the usage of flow cytometry[17, 109-111]. Nevertheless, in all of these methods the cells are not viable, as they are either stained or rely on immunofluorescent detection, requiring cell fixation.

An alternative method, which was introduced in 2008 by Sakaue-Sawano *et al*.[17] revolutionized the cell cycle analysis, as it allows to monitor the cell cycle phases in living cells[19]. The Fluorescent Ubiquitination-based Cell Cycle Indicator (FUCCI) allows to distinguish different cell cycle phases using a dual-color imaging system[19] (Figure 5 B). It takes advantage of the opposing effects of the licensing factor Cdt1 (chromatin Licensing and DNA Replication Factor 1) and its inhibitor Geminin[19]. In eukaryotic cells, Cdt1 accumulates only during G1 phase and prevents that DNA is replicated more than once during the cell cycle. It is degraded in the absence of Geminin, through ubiquitin-dependent proteolysis during the S-phase of the cell cycle. Geminin binds to Cdt1 and prevents MCM (mini chromosome maintenance) chromatin loading. It starts accumulating when Cdt1 levels decrease during S-phase and G2 phase[112]. Geminin levels are low during late mitosis and G1 phase. In the FUCCI assay, each of these proteins is fused to a different fluorescent protein[113]. Usually, Ctd1 is tagged to red fluorescent protein (RFP) and Geminin is tagged to green fluorescent protein (GFP)[114]. When either of the proteins is targeted for degradation upon entry into a specific cell cycle phase, the respective other fluorescent tag is detectable. Cells in S/G2/M are labeled green as GFP-tagged Geminin is expressed, cells in G1 are labeled red as a result of the presence of the RFP-tagged Ctd1 and all cells in the transition state between G1 and S phase, show both GFP and RFP-labeled cells, appearing as yellow. During a short period between M and G1 phase, a small fraction of non-labeled cells are appearing[19, 113, 114].
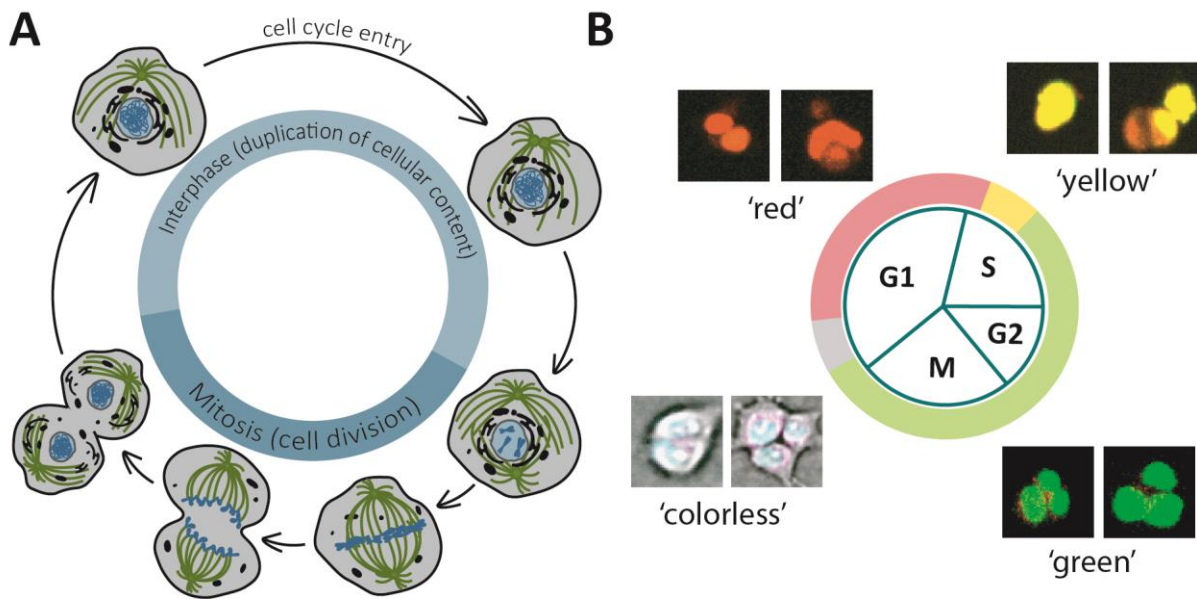
Figure 5: General introduction into the cell cycle and the FUCCI system. **A.** Cell cycle scheme. Figure adapted from Matthews *et al.*[102]. The Interphase represents the longest cell cycle phase. It is divided in three sub-phases (G1-phase, S-phase and G2-phase). Here, the cell prepares itself for cell division. Afterwards the cells are dividing during mitosis into two new daughter cells. **B.** The FUCCI-system allows to distinguish the different cell cycle phases using a dual-color imaging system. Cells in S/G2/M will appear green as GFP-tagged Geminin is expressed, cells in G1 will appear red as a result of the presence of the RFP-tagged Ctd1. Cells in the transition state between G1 and S phase, show both GFP and RFP-labeled cells, and therefore appear yellow. Cell of the period between M and G1 phase are colorless. This figure was created using Adobe Illustrator v26.5.

The FUCCI system has the ability to detect the different cell cycle phases in living cells, which sets it apart from the traditionally used methods[19]. The dynamic changes of the cell cycle progression over a long time-period can be detected in a non-invasive way[115]. The markers are expressed within the cells and no external substances such as dyes, that can harm the cells are used. Also, the dual-color system is easy to visibly observe. The FUCCI system is based on single cell analysis which means that also the heterogeneity of the cells can be taken into account[116].

3.2.6. Impacting cell states by affecting specific cellular compartments

In general, any impact on cell health can influence numerous cellular functions and can also significantly affect cellular compartments such as microtubules or mitochondria. However, this also means that affecting these compartment may result in cell death.

Mitochondria are responsible for producing the majority of the cellular energy in form of adenosine triphosphate (ATP) through oxidative phosphorylation and contribute to many

important cell mechanisms such as different cell death forms. Mitochondrial changes play an essential role in apoptosis induction, and other mechanisms. The mitochondrial outer membrane permeabilization (MOMP) contributes to the release of soluble proteins, leading to apoptotic cell death[85]. Mitochondrial membranes are highly dynamic structures undergoing structural changes every 10 minutes[117, 118]. Mitochondria are also the main source of reactive oxygen species (ROS), which are formed during the energy production process as byproducts. If ROS production becomes excessive due to the influence of certain compounds or an imbalance in cellular antioxidant defenses, it can result in oxidative stress, which can be harmful to the cells. Some substances lead to cell damage by elevating ROS levels in cells. It is crucial to understand the interplay between a compound and ROS production to minimize any interference with cellular energy production and homeostasis[15].

Next to mitochondria, another fundamental structure of eukaryotic cells are microtubules. Microtubules serve as dynamic organizations containing of α- and β-tubulin subunits and play an integral role in maintaining cell shape, facilitating intracellular transport. They are key components of the mitotic spindle during cell division[14]. Due to their central role in many cellular processes, microtubules have garnered a significant interest in the drug development field. Especially in cancer research, microtubule modulating compounds have been extensively employed using their ability to inhibit tumor growth. There is thus a possibility of misinterpretation presumed target-specific effects due to unspecific or unrecognized tubulin binding of a chemical substance. One example to demonstrate the importance of the validation on tubulin binding is, is the compound TH588, which was published to be an mut-T homolog-1 (MTH1) inhibitor by Gad et al. in 2014[119]. The compound showed promising cytotoxic effects in a wide range of human cancer cell lines and even advanced to preclinical studies as a promising candidate for cancer therapy. However, Gul et al.[120] proposed in 2019, that the underlying mechanism for the compounds antiproliferative effect was its ability to modulate microtubules rather than targeting MTH1[120]. Firstly, they found, that the concentration required for the anti-cancer effect was higher than needed for MTH1 inhibition. Secondly, when they performed a CRISPR-mediated knockout of MTH1, they couldn't observe the same effect in cells, indicating that MTH1 is not responsible for the observed phenotype. Furthermore, the compound displayed a concentration-dependent effect on centrosome separation, prolonged mitotic duration, and mitotic slippage[120]. These finding showed the

importance of identification of tubulin-binding to avoid misdirected interpretations of drug-target effects. It also highlights the need for comprehensive studies to uncover the precise molecular mechanism of action of compounds.

3.2.7. Phospholipidosis

Sometimes, it remains uncertain whether an observed phenotypic effect is attributed to a specific target or if the structure of the compound and its physicochemical properties are responsible for the resulting phenotypic response. Here, the phenomenon called phospholipidosis serves as an example.

Phospholipidosis (PL) is a condition characterized by an accumulation of phospholipids in cells and tissues, particularly in lysosomes[21]. Phospholipids are essential components of cell membranes and play a critical role in maintaining the integrity and function of cells. However, in certain circumstances, such as exposure to specific drugs or chemicals, phospholipids can accumulate excessively within the lysosomes, leading to the formation of characteristic lamellar bodies[121]. This occurrence is called drug-induced phopsholipidosis (DIPL) and can result in the disruption of normal lysosomal function, leading to cellular damage and potentially contribution to the development of various diseases. DIPL is a significant concern in drug development and toxicology, as it can interfere with the efficacy and safety of therapeutic agents and chemical compounds. The mechanisms underlying phospholipidosis are complex and not fully understood. There is an ongoing debate whether phospholipidosis is a target-specific process, or a result of non-specific effects caused by the chemical properties of small molecules. One proposed mechanism is that an accumulation of small molecules in lysosomal membranes can trigger the release of lysosomal hydrolases and sphingolipid activator proteins from the surface of intralysosomal luminal vesicles[121]. Another proposed mechanism emphasizes an interference with normal trafficking of lysosomal enzymes, resulting in an accumulation of phospholipids and other cellular components. Furthermore, several targets have been suggested as potential DIPL mediators, including lysosome-specific phospholipase A2 (PLA2G15)[122], or the sphingomyelinase (Asm)-ceraminde system[123]. However, as of now, there are still uncertainties as no single target has been able to fully explain all observed DIPL cases[124]. Regularly, DIPL is linked to structural characteristics of small molecules, especially to cationic amphiphilic drugs (CADs). They contain hydrophobic and hydrophilic domains and carry a positively charged moiety at physiological pH[121]. These

chemical properties provide small molecules with the capacity to passively diffuse through lipid bilayers of lysosomes, in which they will become ionized and trapped due to the acidic pH values[125, 126]. Hence, it is generally accepted that this condition is associated with disturbances in the normal catabolism and turnover of phospholipids.

3.3. Biological Reporters used to assess different cellular processes

Cellular responses to a chemical compound are diverse and complex, which is why it is important when annotating compounds to assess a large number of diverse biological effects, ideally in different cell systems. Next to orthogonal or plate-reader based systems, the use of biological reporters, also known as biomarkers or biosensors, is a good alternative[127]. Biological reporters are molecular entities that can be used to assess and report diverse cell states by interacting with cellular compartments or proteins and translate this binding into a detectable signal such as fluorescence, chemiluminescence or colour-based reactions[127]. Biological reporters encompass different tools, ranging from genetically encoded fluorescent proteins such as GFP or RFP, Luciferase Reporter Assays, pH-sensible indicators to small molecules with light-exciting capabilities.
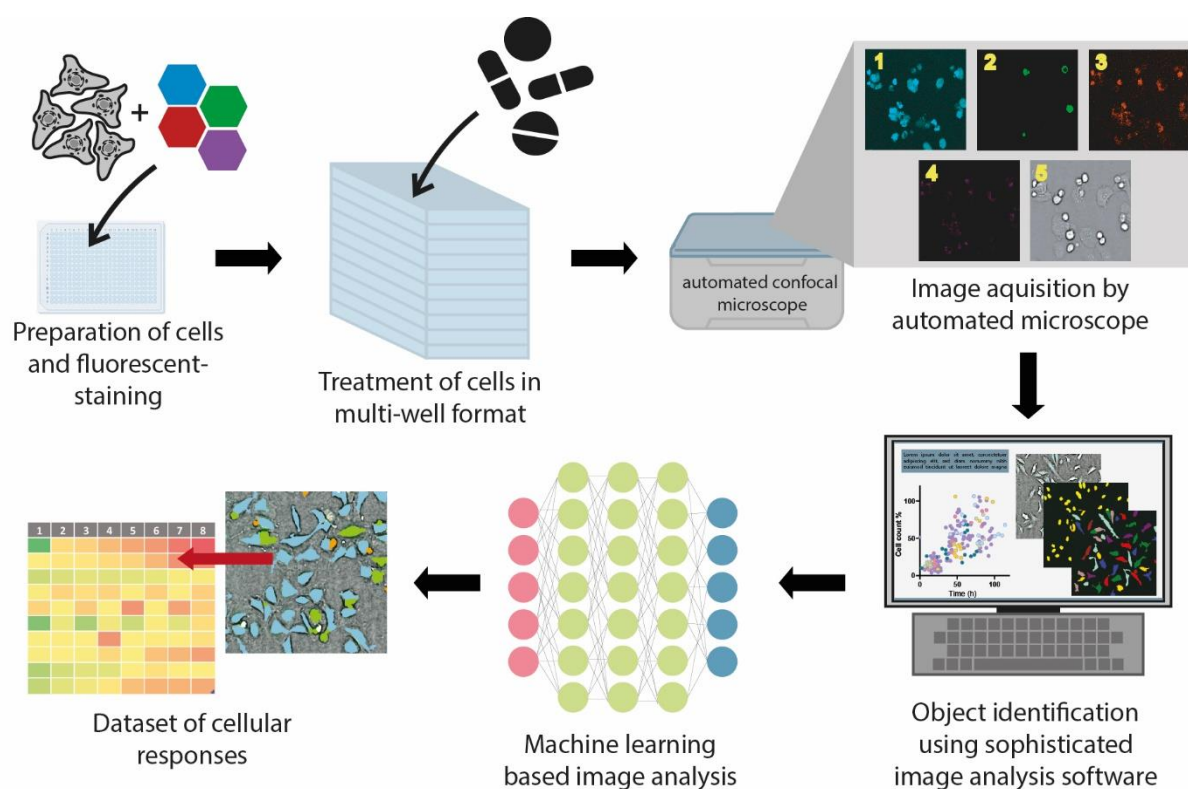
One group of biological reporters are fluorescent staining dyes that can be applied to diverse cell systems. By using fluorescent staining of cells, different cellular mechanisms can be visualized and cellular processes can be quantified with exceptional precision. As of today, a wide array of organelle-staining fluorescent dyes are commercially available[128]. The choice of the fluorescent dye depends on the cellular structure or process of interest, but also the compatibility with the organism and imaging equipment. However, when using several dyes, a major problem for imaging in different channels is so called bleed through or spectral overlap of the different dyes as they often have broad excitation and emission spectra[129]. This should always be kept in mind, when setting up multiple staining protocols. A versatile tool to image specific molecules of interest is the use of small-molecule fluorescent probes. These should not be confused with selective chemical probes mentioned earlier, as they are highly selective and specific probes for cellular imaging, which can be easily used in diverse set ups[128, 130]. They act via different mechanisms, interacting with biomolecules, accumulating via transporters, or participating in metabolism[131]. For example, MitoTracker™ probes have let to major insights,

visualizing mitochondria in diverse cellular systems[128]. MitoTracker™ probes are commercially available with different fluorescent tags for example MitoTracker™ red, passively diffuse across the plasma membranes and accumulates in mitochondria or MitoTracker™ green, which is membrane potential insensitive and represents the mitochondrial mass. They contain an additional reactive moiety such as a benzyl chloride group and cationic properties, which make them cell permeant and mitochondria selective, resulting in their high immobilization in mitochondria[132]. As they are often easily adaptable and applicable for live-cell imaging, they can be used for high-content imaging.

3.3.1. High-content imaging

High-content imaging (HCI) or alternatives such as high content analysis (HCA) or high content screening (HCS) are used to describe automated microscope-based assays, to assess the biological activity in cells or organisms after treatment with thousands of agents such as compounds or siRNAs, in a multi-well format[129, 133]. Typically, the cells or organelles are stained with one or more fluorescent dyes and imaged to detect multiple characteristics (Figure 6). The aim is to identify and measure essential features from the captured image under a particular treatment condition in a short and effective manner[133]. For time- and cost-effectiveness, this can be done simultaneously, using a fast and simple way that provides as much information as possible. Hence, image acquisition is performed using automated microscopes such as confocal microscopes with lower magnification ranges, that allow to capture a larger number of cells for statistically meaningfulness[129].

To extract information from this "*high content*" set-up, modern, fast and sophisticated image-analysis software is used. An image generated by a confocal microscope contains a large amount of data. Depending on the experimental setup and the analysis software used, the output will differ from a basic object identification to more sophisticated *machine learning* based validations up to morphological profiling using deep learning. Typically, first a threshold is applied to the images to identify all objects different from the background. Depending on the cell staining dyes used, individual objects can be detected as either part of the main objects e.g. nucleus as part of the cell body or as an individual object for example dust in the well.

Figure 6: Example for a high content imaging workflow. Starting with the preparation of the cells, including staining of the cells using fluorescent biological reporters and seeding them on a multi-well plate. Afterwards, the cells are treated with compounds of interest and incubated over a distinct time point *e.g.,* 24 hours. Image acquisition is performed using an automated microscope detecting multiple different channels. Images are then analyzed using sophisticated image analysis software that detects different objects, parting them from the background or confounding factors such as dust in the well. Each cell is the profiled using an implemented *machine learning* based algorithm that was trained using different reference compounds. Dataset is generated and then evaluated based on background information such as target potency and selectivity. This figure was created using Adobe Illustrator v26.5.

3.3.2. Live-cell imaging as a versatile tool to detect time-dependent cellular changes

A significant advantage of fluorescent probes is their usability in living cells. The choice between fixed samples and live cells involves balancing the impact and potential damage of the cells with a high signal-to-noise ratio[134]. Fixed cells can be exposed to a much higher light capacity than living cells. Additionally, when using living cells, environmental conditions such as temperature, humidity and CO_2-levels have to be kept constant[134]. Therefore, one has to determine if an assay is amenable to live cell imaging conditions. A time-course study on the environmental conditions as well as the fluorescent probe influence on cell viability should be tested beforehand. Also, photobleaching or phototoxicity of cells should be kept in mind, particularly in the presence of fluorophores, as they can generate free radicals[129, 134].

Nevertheless, especially for compound annotations, live-cell imaging is desired, as dynamic changes and fluctuations can be crucial for the understanding of the cellular response. Monitoring kinetic changes also allows to capture the full compound effect starting from early responses to a potential recovery of the cells. It also helps to verify the optimal time point for the usage of the compound, as off-target effects can occur after extended exposure. This is also crucial for follow up validation of the compound.

3.4. Objectives and Outlook

In this study, various cellular responses to chemical compounds were explored using high-content imaging techniques in living cells. We established a multiplex staining method, which is feasible to be used in live-cells to capture dynamic cellular effects without harming the cells. In an easy and fast way, this method can be used to detect the influence of compounds on cell viability, capture diverse forms of cell death and determine any cell cycle effects. The assay is adaptable to different cell lines and can be used in combination with other assays to assess effects, such as mitochondrial health, changes in cytoskeletal morphology or phospholipidosis (Figure 7). We thus were able to annotate newly developed kinases inhibitors and investigate their effect on different cell states.

The establishment and suitability of the assay was tested on a small compound library containing 215 compounds in three different cell lines (section 4.1.1. and Appendix A). To allow diverse users to reproduce the assay and to make it accessible, we published an exceptionally detailed protocol (section 4.1.2. and Appendix B).

As part of a EUbOPEN, I contributed to generate and annotate the largest freely available set of high-quality chemical modulators targeting the human proteome. Within this innovative medicine initiative (IMI)-funded project, I evaluated over 1000 diverse chemical compounds for their effects on cellular quality, performing a two-step viability assessment. To ensure reproducibility and visibility of the data generated, both steps of this characterization pipeline have been described and published as step-bye-step protocols in the book "*Chemogenomics*" (section 4.2. and Appendix C and Appendix D). Additionally, all data generated during that process was made available to the public to provide a comprehensive understanding of the cellular quality of the compounds (section 4.2.).

Using the assay, I also validated the impact of 150 selective chemical probes and controls provided by the donated chemical probes (DCP) set regarding their cell viability and phenotypic characteristics across different standard cancer cell lines (section 4.3.1. and Appendix E). This annotation in model cell lines was used as a basis to compare the effect of these probes to sophisticated human organoids of colorectal cancer. In other projects, I characterized off-target activity of different chemical probes such as the macrocyclic scaffold compound CK156, which selectively targets Serine/threonine kinase 17A (STK17A) (section 4.3.2. and Appendix F) or the dual Discoidin domain receptor (DDR)/p38 chemical probe, SR-302 (section 4.3.2. and Appendix G). Additionally, I evaluated the effect of Cyclin-dependent kinase 16 (CDK16) inhibitors on the cell cycle, by assaying different compounds targeting the PCTAIRE kinase family (section 4.4. and Appendix H).

Lastly, I tested two diverse compound sets in an optimized assay, enabling the evaluation of time-dependent phospholipidosis (PL) induction. Within this collaboration, I analyzed two training sets, containing a total of 290 compounds, which were used to develop a precise *machine learning* (ML) algorithm for the prediction of PL inducers. Leveraging the algorithm's capabilities, we made exciting discoveries regarding the structural influence on PL. On the other hand, by evaluating different chemical probes from the chemical probes portal and their matching negative control, we found that closely related molecules, can differ in their ability to induce PL (section 4.5. and Appendix I).
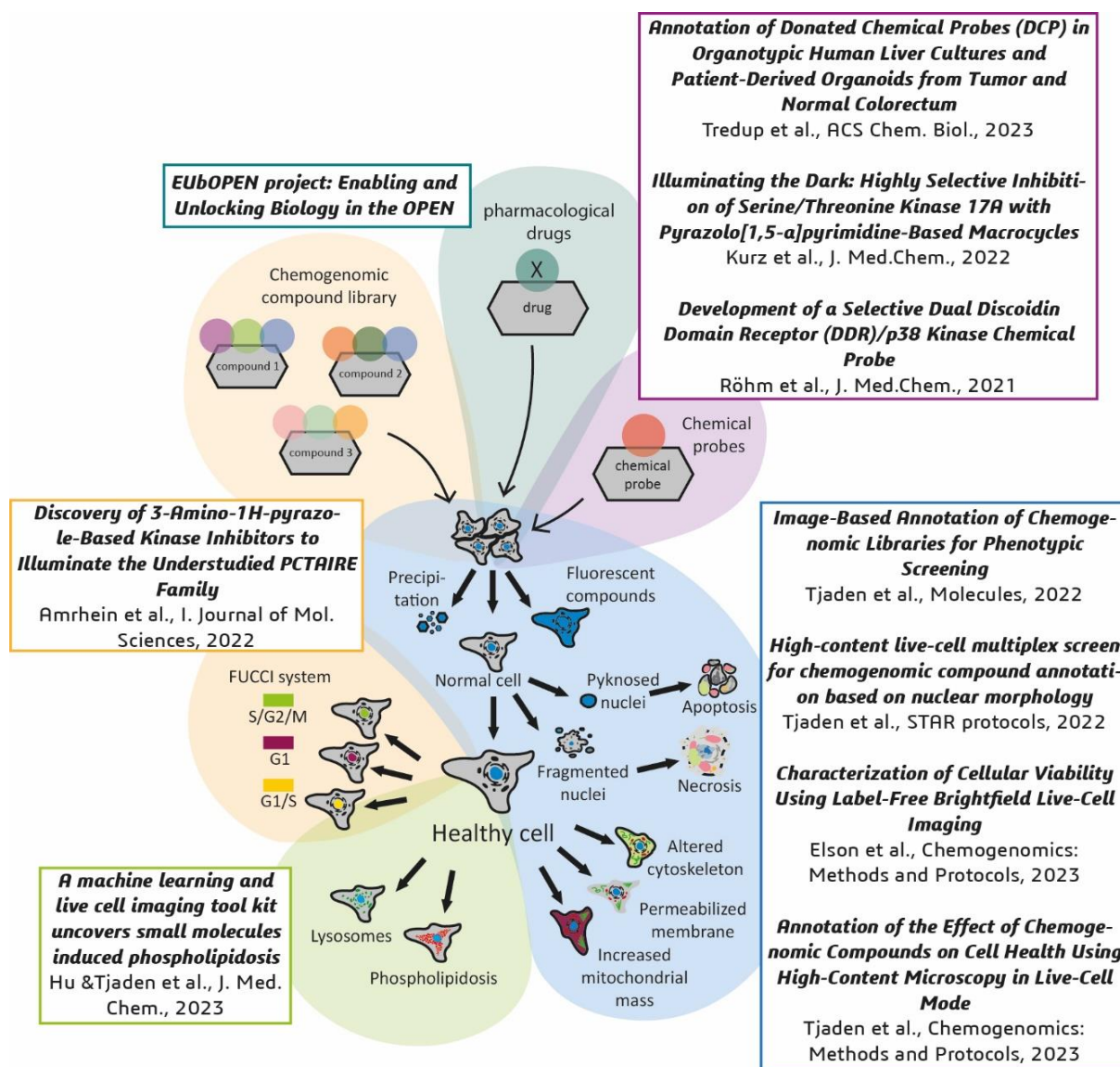
Figure 7: Graphical illustration of the connection between the here presented publications. This figure was created using Adobe Illustrator v26.5.

4. Results

4.1. Compound annotation in living cells using high-content imaging

As part of the characterization process of small molecules, testing their influence on cell health especially at different concentrations, represents a crucial step. To conduct this cellular quality control efficiently and to generate a large set of data about a compound's influence on different cell states and cellular responses, high-content imaging is a suitable method.

4.1.1. Development of a Multiplex high-content imaging approach

I developed a high-content assays platform to characterize compounds in a multi-dimensional way compounds for their cellular response in live-cells (Appendix A), which will be presented in the following section.

"Image-Based Annotation of Chemogenomic Libraries for Phenotypic Screening"[114]

In this work I developed a high-content assay in living cells to annotate compound libraries. The assay was originally based on a single time point protocol called "High-Via", published by Howarth *et al*. in 2020[74], which described an image-based detection of different cell death forms. In comparison to the High-Via protocol, where the cells are categorized into 4 different groups ("healthy", "early/late apoptotic", "necrotic" and "lysed"), we observed that the nuclear phenotype alone, when categorized into "healthy nuclei", "fragmented nuclei" or "pyknosed nuclei", resulted in similar cytotoxicity profiles, compared to the overall phenotype (Figure 8 A). To identify cell nuclei, the DNA-staining dye Hoechst33342 was used. This allowed us to make an equivalent statement about cell viability, using only one of the five distinct channels of the microscope.

Relying solely on a single channel, however, increased the risk of assay interference from compounds sharing similar fluorescent characteristics or from poorly soluble compounds, which tend to produce a significant fluorescent background. Hence, I introduced a preliminary categorization in "Hoechst High Intensity objects". These objects correspond to either a precipitation or auto fluorescent activity of a compound (Figure 8 B and C). With four other channels available, the assay was now expandable and could be adapted to investigate diverse cellular responses. I evaluated and optimized the protocol to be used in different cell lines. The

main cell lines used in this publication were human embryonic kidney cells (HEK293T), osteosarcoma cells (U2OS) and untransformed human fibroblasts (MRC-9).

To identify next to primary effects also secondary cellular changes, I adapted the protocol for a continuous usage over 72 hours. When conducting live-cell imaging, it is essential to employ the optimal dye concentration. This concentration should strike a balance between the dye not interfering with cellular functions while at the same time providing a sufficient fluorescent signal for a robust detection. I therefore validated the influence of the cell staining dyes on their effect on cell viability using AlamarBlue™ (Figure 8 D).
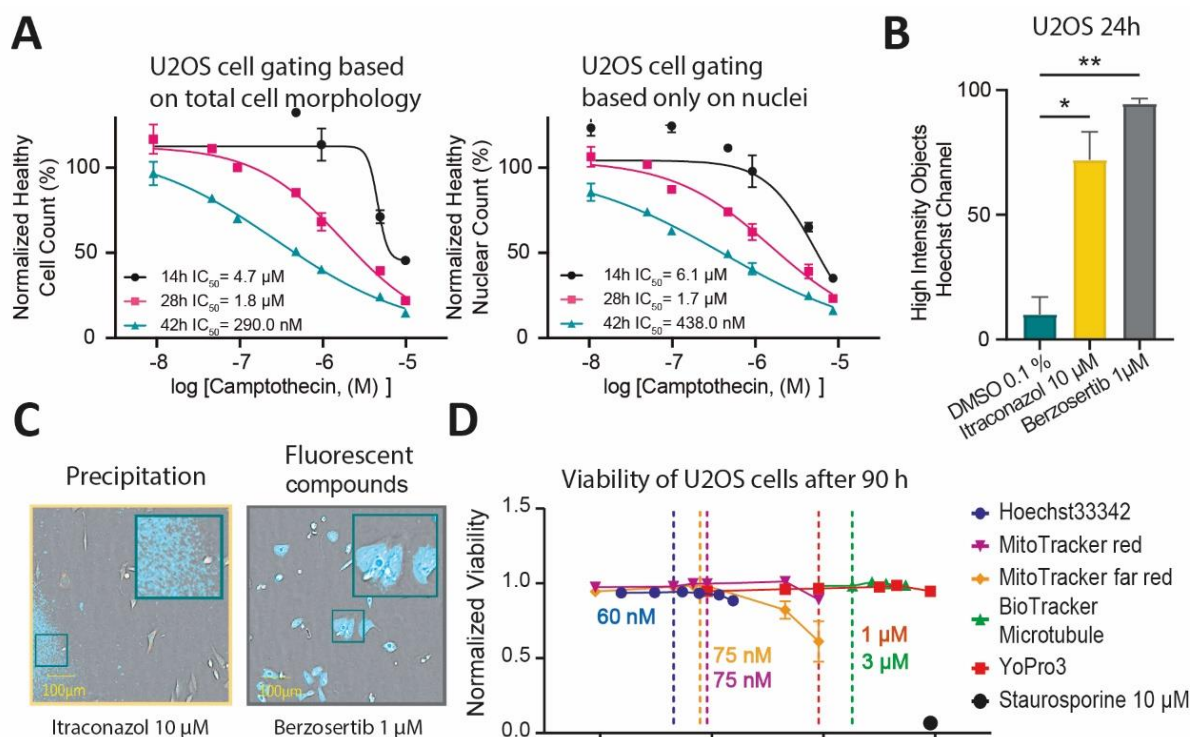

Figure 8: Viability assessment using nuclear morphology in comparison to overall cell morphology. This figure is adjusted from Figure 1 and 2 of Tjaden *et al.*[114]. **A.** Normalized healthy cell count in comparison to normalized healthy nuclear count of U2OS cells exposed to different concentrations (0.01 µM, 0.05 µM, 0.1 µM, 0.5 µM, 1 µM, 5 µM, 10 µM) of camptothecin with calculated IC50 values after 14h, 28h and 42h. **B.** Hoechst High Intensity objects after 24h of compound exposure (itraconazol 10 µM, berzosertib 1 µM) to U2OS cells in comparison to cells exposed to DMSO 0.1 %. Error bars show SEM of three technical replicates. Statistical evaluation shows two-way ANOVA analysis calculated using Graph Pad Prism 8.4.3. (*=0.0332, **=0.0021). **C.** Brightfield confocal image of stained (blue: DNA/nuclei) U2OS cells after 24h of compound exposure (itraconazol 10 µM, berzosertib 1 µM). Precipitation of 10 µM itraconazol and fluorescence of 1 µM berzosertib exposure shown as Hoechst High Intensity objects are highlighted. **D.** AlamarBlue™ viability assessment of U2OS

cells. Graphs show normalized viability after 90h of exposure to different cell staining dyes at six different concentrations in comparison to cells exposed to 10 µM of staurosporine. Vertical lines show optimal concentration used in subsequent assays (Hoechst33342 60 nM, MitoTracker™ red 75 nM, MitoTracker™ far red 75 nM, YoPro3 1 µM, BioTracker™ Microtubul 3 µM). Error bars show SEM of four technical replicates.

In the publication, the assay was expanded, to detect in addition to diverse cell death forms, additional effects on membrane integrity, mitochondrial health and influences on tubulin integrity. Therefore, the cells were additionally stained with MitoTracker™ red to investigate changes in mitochondrial mass, the taxol-derived tubulin dye BioTracker™488 Green to observe microtubule effects, as well as Annexin V to identify apoptosis. This expanded protocol was subsequently verified by testing 21 reference compounds, each having well documented cellular responses. These references were also used to establish criteria for compound flagging, determining their suitability to be included in a chemical tool compound library. A workflow of the resulting assay can be found in Figure 9. This assay structure will be further referred to as the "*Multiplex assay*". The applicability of the assay in a medium throughput format was demonstrated by testing a small library, comprising 215 chemogenomic compounds. The compounds were tested at two different concentrations (1 µM and 10 µM) in three different cell lines (HEK293T, U2OS, MRC-9) over 24 h.
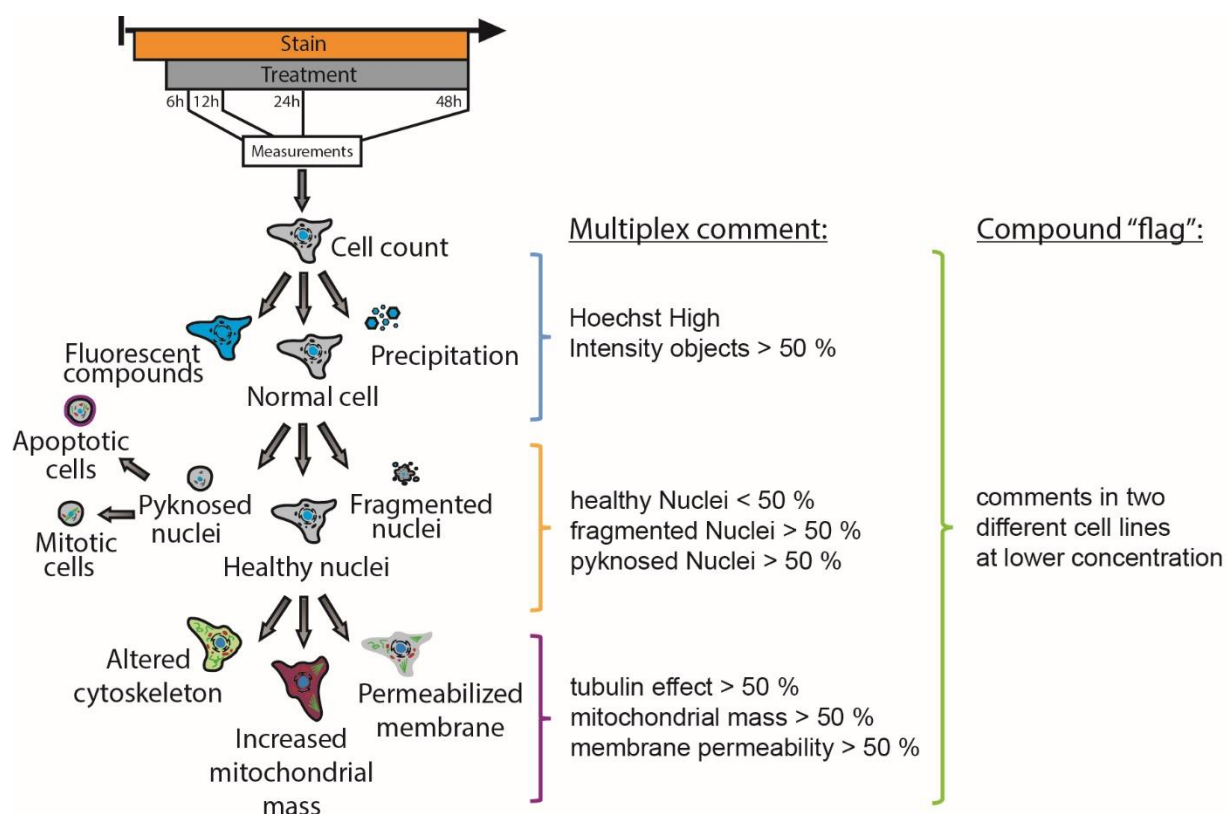
Figure 9: *Multiplex* **assay tree with property thresholds.** This figure is adjusted from Figure 4 of Tjaden *et al*.[114]. Tree-based workflow of *Multiplex* assay protocol showing possible comments for compound annotation and the criteria for a compound to be "*flagged*".

To show the adaptability of the *Multiplex* assay protocol, I also tested the protocol in combination with the fluorescent ubiquitination-based cell cycle indicator (FUCCI), that allows to distinguish different cell cycle phases using a dual-color imaging system. Within this adaptation of the protocol, the effects of 18 different test compounds in human colorectal cancer cells (HCT116) was examined and the influence on both cell viability as well as the cell cycle over 70h was validated.

4.1.2. Step by step protocol to improve data availability and reproducibility

In order to make the developed assay accessible to the community and provide all necessary information to reproduce our results, we published a detailed step-by-step protocol (Appendix B). In the following section, I want to elaborate on the significant insights that can be gained by using the protocol, how it facilitates the reproducibility of the *Multiplex* assay described above and the advantages of sharing such a protocol with the scientific community.

"High-Content Live-Cell Multiplex Screen for Chemogenomic Compound Annotation based on Nuclear Morphology"[135]

The protocol describes the workflow to test 135 test compounds in the *Multiplex* assay at two different concentrations in U2OS cells on one 384 well-microplate. First there is an elaboration on all measures and procedures that need to be taken before the start of the experiment. It showes a detailed description of the preparation of the cells, as well as the compounds and how to set up all instruments that will be used. The user can also find a table indicating all reagents that were used and where they have been sourced from. A large portion of the protocol is allocated to the analysis of the images. The analysis was performed applying the CellPathfinder software from Yokogawa. A general workflow is shown in Figure 10 A. Here, first we defined two objects called "*Cellbody*" and "*Nucleus*" (Figure 10 B). The Nuclei are defined to be included in the cellbodies, one nucleus per cell. Afterwards a *machine learning* based algorithm, implemented in the software, was trained to gate all cells into different categories (Figure 10 C). The categories were based on the tree principle, shown in Figure 9. All analysis files and additional information were included in the protocol, ready to use, to ensure reproducibility and easy access to the methods.

To train the algorithm, a test set comprising 21 compounds was used[114]. The protocol recommends to always add 6 of these reference compounds to every experimental setup, to assess whether the algorithm performs as desired and yields statistically significant results. Here, staurosporine is proposed for apoptotic cell death and pyknosed nuclei, paclitaxel to indicate structural changes of tubulin, milciclib for an increase in mitochondrial mass, daunorubicin to train the software on fragmented nuclei, apoptotic cells and an increase in mitochondrial mass, digitonin as a detergent for cell membrane permeabilization and berzosertib for Hoechst High Intensity objects.
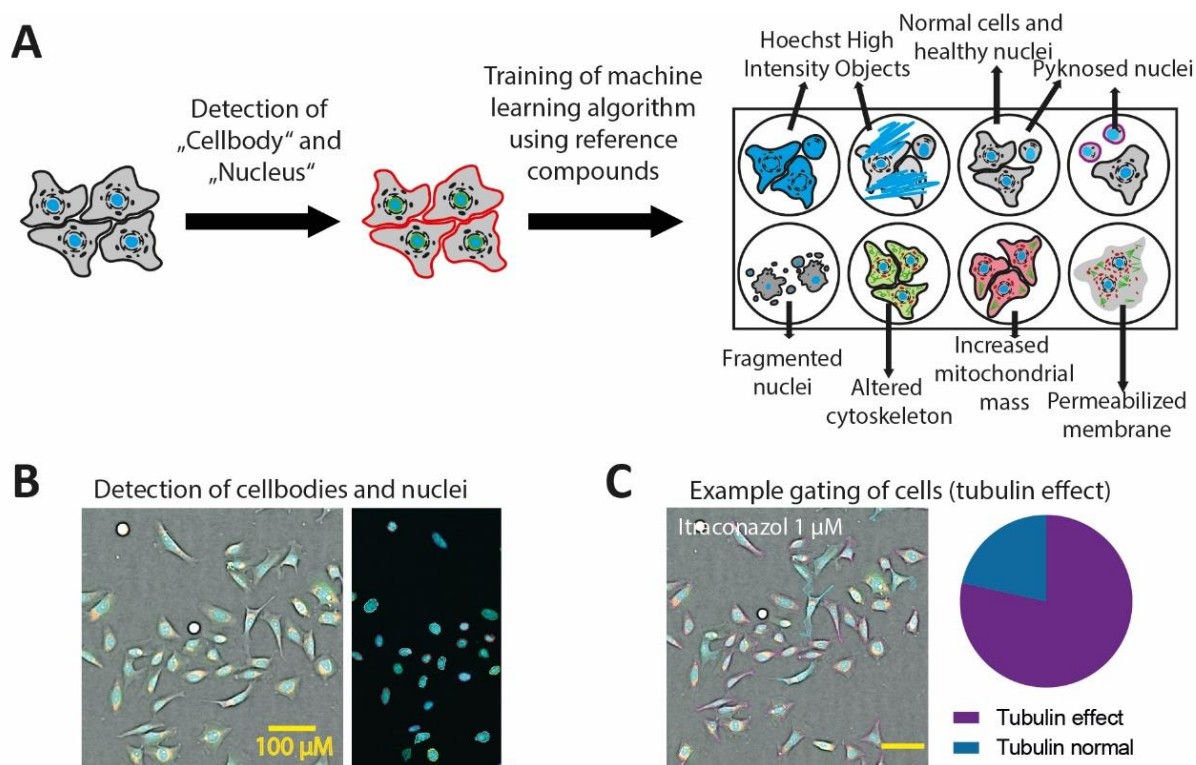
Figure 10: *Multiplex* assay analysis using CellPathfinder software. This figure is adjusted from figure 6 and figure 9 of Tjaden *et al*.[135]. **A.** Scheme of the general workflow of *machine learning* based analysis. First detection of "*Cellbody*" und "*Nucleus*", then training of *machine learning* based algorithm using distinct references. **B.** Processed image of stained (blue: DNA/nuclei, green: microtubule, red: mitochondria content, magenta: Annexin V apoptosis marker) U2OS cells after 24h of exposure to Itraconazol (1 µM) and detection of nuclei and cell bodies after analysis using the CellPathfinder software. **C.** Gating of U2OS cells in tubulin effect or tubulin normal using the CellPathfinder analysis after exposure to 1 µM of Itraconazol after 24h. Pie charts show the ratios of the different gating steps.

To further help scientific progress and enabling drug development, an open access to data is desirable. Particularly, a high amount of diverse data and easy data accessibility can significantly improve compound characterization. Therefore, the protocol shows additionally to the analysis and evaluation of the data, how the data can be made readily available to the entire scientific community in a straightforward manner. It explains a detailed process to swiftly and easily upload the image data into a public database, along with the essential information relevant to the experiment. It describes how to deposit all acquired images for the compound annotation on the European Molecular Biology Laboratory (EMBL) European Bioinformatics Institute's (EMBL-EBI) repository website BioImage Archive (https://www.ebi.ac.uk/bioimage-archive/).

At the end of the protocol, there is also a detailed troubleshooting section, where comprehensive explanations of various problems, that may arise during the execution of the protocol, are described, alongside with multiple solutions to address them.

4.2. A versatile pipeline to characterize compounds on their cellular quality within the EUbOPEN project

The Europe-wide project called EUbOPEN (Enabling and Unlocking Biology in the OPEN), is an innovative medicine initiative (IMI) funded project, which aims to generate the largest freely available set of high quality chemical modulators for human proteins[49]. All compounds within this set have to undergo comprehensive profiling, including assessments of their selectivity, potency and cellular activity. One integral aspect of the project is the development of novel technologies to discover and characterize compounds. For this, within our working-group we implemented a two-step pipeline to assess both cell viability as well as the overall cellular quality of compounds. Both steps of this versatile pipeline were reported in the following publications (Appendix C and D). In the subsequent section, I will provide a brief overview of the key steps involved in the characterization process using this pipeline.

"Characterization of Cellular Viability Using Label-Free Brightfield Live-Cell Imaging"[136] and "Annotation of the Effect of Chemogenomic Compounds on Cell Health Using High-Content Microscopy in Live-Cell Mode"[137]

In the established workflow, all compounds that were potential candidates for inclusion in the set of high quality chemical modulators of the EUbOPEN project, underwent a comprehensive two-step cellular quality assessment. First, the general influence on viability of the compounds was tested using a label-free brightfield live-cell imaging approach (Appendix C). Here the growth rate and phenotypic appearances were assessed. Noteworthy compounds, identified in the first screen, were subsequently tested in the second screen (Appendix D). This involved the *Multiplex* assay, a high-content live-cell imaging approach, which has been explained in section 4.1. and 4.2. Following this, compounds that showed specific threshold values of the different phenotypic characteristics were evaluated based on their specific targets, leading to either their inclusion in the set or a compound flag (see Figure 9). Moreover, all data generated

through this process were made accessible to the public via the BioImage Archive as described by Tjaden *et al.*[135].

The initial evaluation step has been explained by Elson *et al.*[136] (Appendix C). In this report, a brightfield live-cell imaging approach using the IncuCyte (Satorius) live cell imaging system has been presented. The protocol described a workflow to test 304 compounds (including controls) in a 384-well format. The compounds were tested at a relatively high concentration (10 μM) in human embryonic kidney cells (HEK293T), osteosarcoma cells (U2OS) and untransformed fibroblasts (MRC-9).

Their influence on cell viability was assessed by validating effects on the confluence using the growth rate. The growth rate (GR) is a normalized parameter, which was introduced by Hafner *et al.*[138] in 2016. It includes the initial cell confluence and is insensitive to the proliferation rate of the cells. It therefore allows the characterization of compounds regarding a partial inhibition, cytostatic properties, cytotoxicity or a healthy outcome (Figure 11). The protocol comprises a detailed description of all procedures, which included cell culture and treatment, image acquisition and analysis and a comprehensive guide on how to calculate the growth rate, illustrated through an example. Compounds with a growth rate smaller than 0.5 were tested in the second *Multiplex* assay. In addition to the growth rate, compounds exhibiting exceptional phenotypes, obtained by visual inspection, were marked for further evaluation.
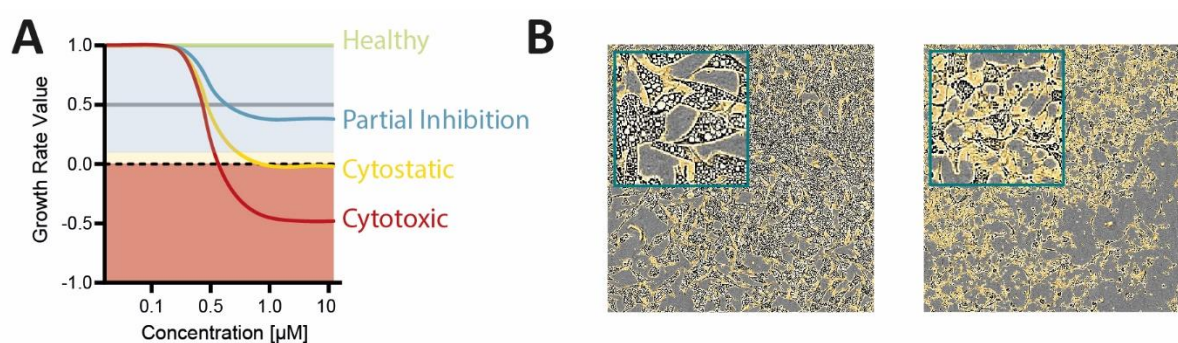
Figure 11: Conditions for testing compound in second *Multiplex* assay. This figure is adjusted from Hafner *et al.*[138] and Elson *et al.*[136]. **A.** Simplified representation of growth rate calculation of healthy, partial inhibition, cytostatic or cytotoxic outcome. **B.** Example images (brightfield, Incucyte®S3) of different phenotypes observed in U2OS cells after confluence analysis using Incucyte®Base analysis software.

The second screening step has been described in Tjaden *et al.*[137] (Appendix D). In this book chapter, the process for evaluating 135 compounds in the *Multiplex* assay has been outlined. They were examined in the same three cell lines (HEK293T, U2OS and MRC-9) and at the same concentration used in the initial primary screening (10 µM), as well as at a concentration that is ten times lower (1 µM). This ready-to-use protocol is an easy to use step-by-step guide to perform the *Multiplex* assay in different cell lines for a distinct purpose.

The whole pipeline aimed to characterize compounds for potential inclusion in a chemogenomic compound set. Furthermore, it served as a means to validate the cellular quality of various diverse compounds across multiple projects. Within the EUbOPEN initiative, over 1000 compounds were characterized so far. The generated characterization data was shared through various open-source databases, including the database of bioactive drug-like small molecules of the EMBL called ChEMBL (https://www.ebi.ac.uk/chembl/) and the EUbOPEN's own database platform gateway (https://gateway.eubopen.org/). All data can be found here: https://www.ebi.ac.uk/biostudies/bioimages/studies/ including the studies #S-BIAD730, #S-BIAD733, #S-BIAD145 and # S-BIAD568.

4.3. Application of the high-content assay called "*Multiplex* assay" to characterize chemical probes

For a compound to be considered as a probe and potentially be included in a probe set, it must meet stringent criteria und undergo comprehensive profiling. One of the requirements that compounds must satisfy is that they should not exhibit cytotoxicity at a concentration of 10 µM, unless it is related to the inhibition of the target. A significant advantage of using chemical probes, is the assurance that the compounds are well-annotated and a link between a phenotype and a target can be made with some confidence. This helps minimizing the possibility of inaccurate results due to unspecific or unknown events. Therefore, the compounds presented in the following were annotated based on their cellular viability and unspecific phenotypic effects using the *Multiplex* assay.

### 4.3.1. Annotation of compounds included in the donated chemical probes library

To date, the donated chemical probes set comprised at 97 chemical probes (https://www.thesgc.org/donated-chemical-probes). It originated in an exceptional project, where a consortium of pharmaceutical companies including AbbVie, Bayer, Boehringer Ingelheim, Janssen, MSD, Pfizer and Takeda have worked together with the Structural Genomics Consortium (SGC) to provide highly selective, potent and broadly characterized tool compounds to the research community[43]. The collaboration not only supplied the compounds with accompanying selectivity and potency data, but also consistently recommended a concentration at which the probe molecule should be used and, if available, provided an inactive control compound[16]. The DCP set comprised well-validated chemical probes for a broad range of targets including protein kinases, G-protein-coupled receptors (GPCR), ion channels and many more. In the following publication (Appendix E), Tredup *et al*. introduced the DCP project and presented the progress and developments of the recent years. The publication showed a variety of *in vitro* and cellular characterization data of the compounds, such as kinome wide profiling from KINOMEScan at DiscoverX (Eurofins)[139] containing 468 kinases or in-cell target engagement data using NanoBRET. The compounds were annotated based on their cellular quality to reduce the risk of inappropriate use and any potential misinterpretation of results. The authors additionally characterized the set in relevant human primary models, such as hepatotoxicity screening in primary human liver spheroids and in patient-derived colorectal cancer organoids and matched normal-adjacent epithelium. In the following section, I want to highlight the cellular annotation of the DCP library using the *Multiplex* assay (Figure 12 A).

"Deep Annotation of Donated Chemical Probes (DCP) in Organotypic Human Liver Cultures and Patient-Derived Organoids from Tumor and Normal Colorectum"[16]

As mentioned above, the DCP collection, comprised highly-selective, openly available tool compounds, that can be used to study a wide range of biological processes and targets. In total, 150 compounds compiling chemical probes and matching negative controls, were tested in three different cell lines (HEK293T, U2OS, MRC-9) at their recommended concentration (1x), as well as at a 10-fold higher (10x) concentration using the *Multiplex* assay. We found, increased probability of unwanted toxic or phenotypic events at the ten-fold higher concentration, highlighting the importance of appropriate dosing. However, at their

recommended concentration, just a few compounds showed unexpected toxicities or cellular effects (Figure 12 B). Only three compounds decreased the healthy cell count in HEK293T cells to less than 40 % comprising a bromodomain (BRD2/3/4/T) targeting compound GSK046, the negative control of the y-secretase complex inhibitor MRK-560 (GSI-NC) and the MALT1 probe compound NVS-MALT1 (Figure 12 C).
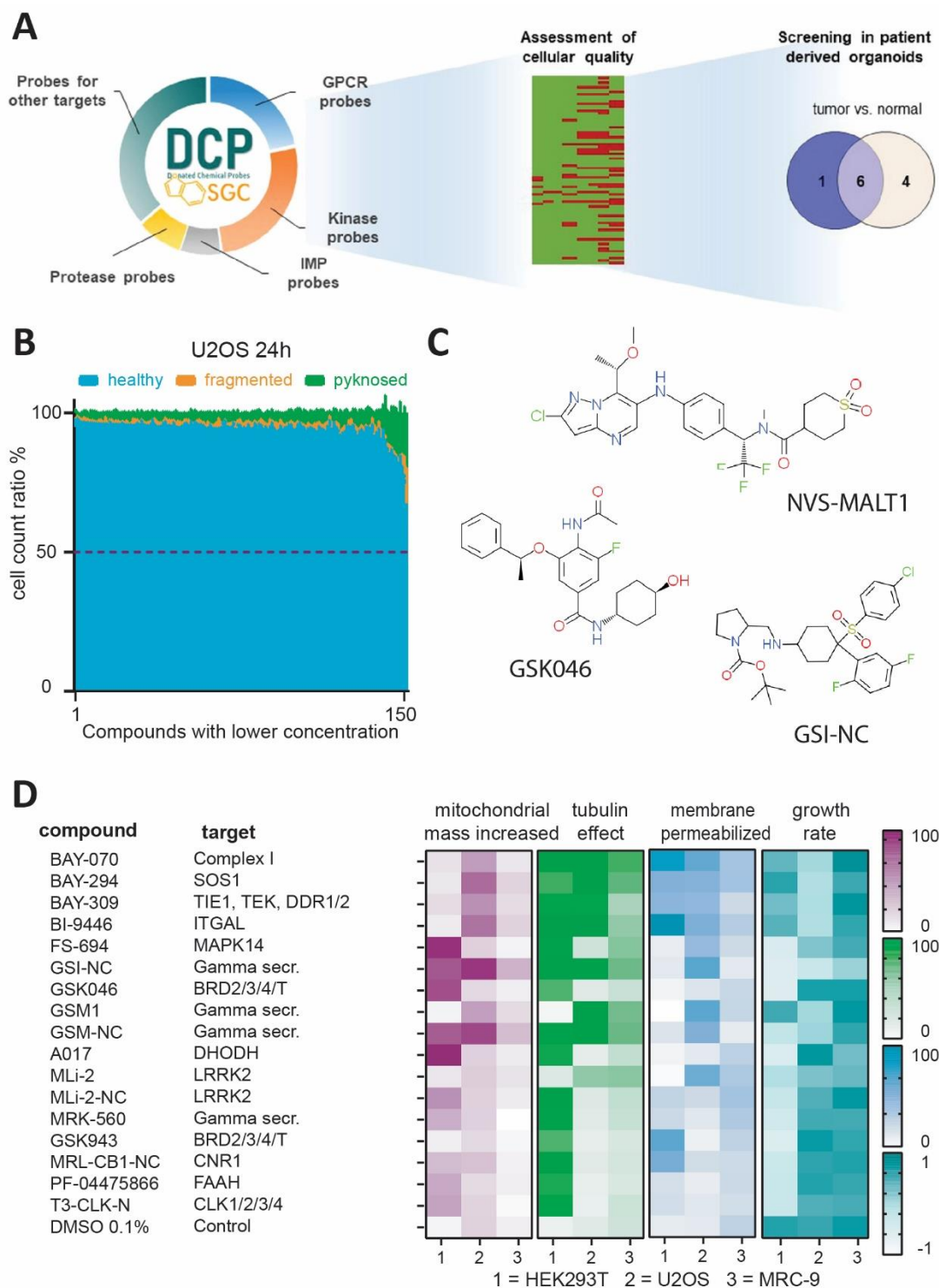
Figure 12: Annotation of Donated Chemical Probes set in cells. This figure is adjusted from Tredup *et al.*[16]. **A.** Overview of the DCP set, which includes 97 chemical probes and display of their target categories. Additionally, outline of the process of annotating them in cells and in primary human organoids. **B.** Cell count ratio of different nuclei gating in U2OS cells after 24h. Compounds shown at their recommended concentration. Healthy nuclei count at 50% (viability threshold) is marked with a red line. **C.** Compound structures of compounds that showed less than 50% of healthy nuclei in HEK293T cells after treatment at recommended concentration (1x). Structures were used from the SGC DCP website (https://www.thesgc.org/donated-chemical-probes) **D.** Heat map of phenotypic properties (increase in mitochondrial mass, tubulin effect, membrane permeabilization) and growth rate values after 24h of "*flagged*" compounds with their distinct target in HEK293T, U2OS and MRC-9. Values are calculated from mean of two biological duplicates.

Furthermore, at the recommended concentration, only 17 compounds showed phenotypic effects higher than 50 % after 24h (Figure 12 D). 70 % of these compounds were negative controls. Negative controls are structurally similar compounds with less or no activity on the target. Often they are less characterized, resulting in a higher probability to cause unspecific or toxic side effects. The five probes that were flagged due to a phenotypic event, were the y-secretase complex probes MRK-560 and GSM1, the BRD2/3/4/T inhibitor GSK046, the MAPK14 inhibitor FS-694 and the LRRK1 inhibitor MLi-2. Based on these results, the authors stressed the importance of using probe compounds at their recommended concentration and highlighted the need to annotate compounds in diverse systems.

4.3.2. Characterisation of chemical probe candidates for the Serine/threonine kinase 17A (STK17A) and the dual DDR/p38 chemical probe

In the following sections I would like to highlight two projects (Appendix F and G) in which we characterized probe candidates, which were evaluated by a team of scientific experts und received approval to be classified as a chemical probe.

"Illuminating the Dark: Highly Selective Inhibition of Serine/Threonine Kinase 17A with Pyrazolo[1,5-a]pyrimidine-Based Macrocycles"[140] (Appendix F)

After many years of ongoing research in the field of biological target identification, and although many proteins have been extensively studied and well characterized, there are still numerous proteins that have not yet been fully explored and are not liganded[141]. Nevertheless, these "*untargeted*" or "*dark*" proteins may play important roles in maintaining cellular functions, but remain undiscovered[141]. In the case of kinases, approximately three quarter of

the human kinome consists of proteins that have not yet been studied in detail[142, 143]. In most cases the identification of these "*dark kinases*" and the determination of their biological function seems to be more challenging and requires comprehensive screening efforts[144]. This is in part due to the conserved ATP-binding site of kinases, which is often the most druggable part of this protein family. Using highly selective tool compounds, targeting dark kinases, can gradually improve the knowledge about these proteins.

For this, Kurz *et al.*[140] developed an exclusively selective inhibitor called CK156 for the dark kinase death-associated protein kinase-related apoptosis-inducing protein kinase 1 (DRAK1) (Figure 13 A). DRAK1 is part of the death-associated protein kinase (DAPK) family which includes also DRAK2, DAPK1, DAPK2 and DAPK3. It is also known as serine/threonine kinase 17A (STK17 A) and regulates many different cellular processes, such as autophagy and apoptosis[145, 146]. As a starting point for the compound CK156, Kurz *et al.* did a structure-based optimization of the casein kinase 2 (CK2) inhibitor IC19, which was first introduced by Krämer *et al.* in 2020[147] (Figure 13 B). They used the promising hinge binding motif, a pyrazolo[1,5-a]pyrimidine structure, present in this compound, for the development of a macrocyclic structure. Macrocyclization can be favorable for alteration of the biological and physiochemical properties and can improve the selectivity in comparison to acyclic analogues[148]. The compound was further optimized by improving the back pocket interactions by introducing space-filling benzylamide derivates. A kinome wide selectivity screen (KINOMEscan) confirmed the remarkable selectivity of CK156 with exclusive affinity for DRAK1. Target engagement was also confirmed in cells using NanoBRET target engagement assay. Additionally, a crystal-structure of DRAK1 revealed a classical type I binding mode of the compound.

To assess the effect of the compound on general cell viability, I performed the high-content *Multiplex* assay which was introduced in section 4.1 and 4.2. in HEK293T, U2OS and MRC-9 cells (Figure 13 C-E). At 10 µM, the compounds showed a decrease in healthy cells (Figure 13 C). The highest impact was observed for U2OS cells, where almost 50 % of the cells showed effects on cell viability. For HEK293T and MRC-9 cells the effect was not as severe (Figure 13 D). At 1 µM, no cytotoxic effects were observed in all three cell lines. Additionally, no notable effect on microtubule or mitochondrial damage was detected at both concentrations (Figure 13 E).

DRAK1 has been described to be highly overexpressed in glioblastoma cells and knock down of this kinase by shRNA resulted in an alteration of cell shape and cell proliferation[149]. To test if an inhibition of DRAK1 results in the same phenotype, the compound was tested in four different glioblastoma cells. However, no correlation between DRAK1 expression and the influence on cell viability or morphological changes was detected.
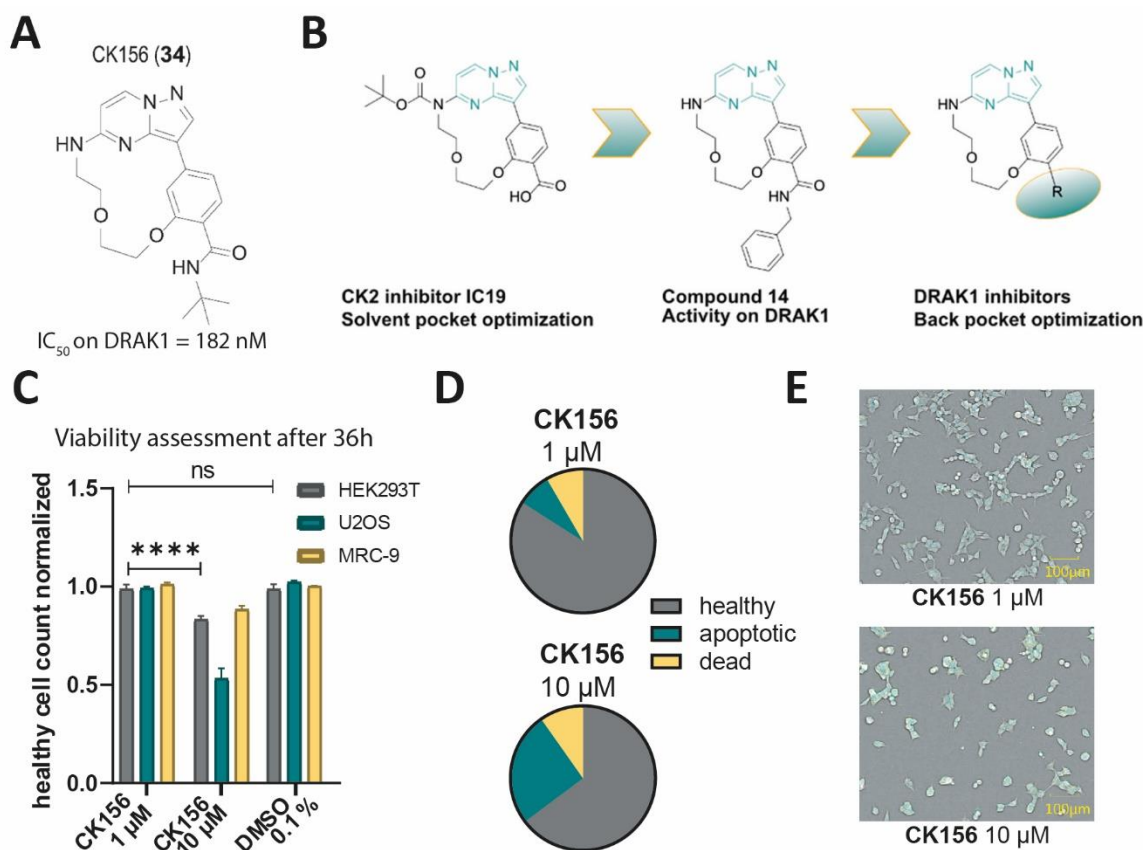
Figure 13: Viability assessment for compound CK156 using *Multiplex* assay. This figure is adjusted from Kurz *et al.*[140]. A. Chemical structure of CK156 with EC_{50} value on DRAK1, which was confirmed using NanoBRET target engagement assay. B. Schematic overview of the development of DRAK1 inhibitors by exploring the back pocket. Optimization of interactions in the solvent pocket resulted in the highly selective CK2 inhibitor IC19. Modification of the back pocket interaction in this study identified the lead structure (14), which lead to selective DRAK1 inhibitors. C. Healthy cell count after 36 h of 10 and 1 µM compound exposure [CK156] normalized to healthy cells exposed to DMSO (0.1%) in HEK293T, U2OS, and MRC-9 cells. Error bars show standard error of the mean of quadruplicates. Significance was calculated using a two-way ANOVA analysis. D. Fraction of healthy, apoptotic and dead cells after 18 h of 10 and 1 µM compound exposure [CK156] in HEK293T cells. E. Brightfield confocal image of stained HEK293T cells after 18 h of 10 and 1 µM compound exposure [CK156] in comparison to 0.1% DMSO control (blue: DNA/nuclei, green: microtubule, red: mitochondria, and magenta: Annexin V apoptosis marker).

"Development of a Selective Dual Discoidin Domain Receptor (DDR)/p38 Kinase Chemical Probe"[150] (Appendix G)

In another project lead by Sandra Röhm, we developed and characterized a cell-active dual inhibitor for the discoidin domain receptors 1 and 2 (DDR1/2) together with p38, which was accepted as a dual chemical probe. DDR kinases are members of the receptor tyrosine kinase (RTK) family and play a pivotal role in fibrotic disorders, atherosclerosis and divers forms of cancer[151]. Röhm *et al.*[150] optimized the recently published p38 mitogen-activated protein kinase inhibitor 7, which originally derived from compound VPC-00628[152], to a series of potent DDR1/2 kinase inhibitors. After extensive characterization using DSF screening and crystallization, their work resulted in a kinome-wide selective (tested against 468 kinases) compound called SR-302 (Figure 14 A). Employing a NanoBRET target engagement assay, the compound displayed EC_{50} values of 0.023 and 0.018 µM for DDR1/2, respectively and 0.125 and 0.196 µM for p38α/β, respectively indicating a high potency of the compounds in cells. Additionally, a negative control compound called SR-301 was proposed and characterized (Figure 14 A). To investigate their dose-and time-dependent effect on cell viability, the probe compound (SR-302) and the negative control (SR-301) were tested in the high-content *Multiplex* assay (see section 4.1 and 4.2) (Figure 14 B-C). At the lower concentration (1 µM), none of the compounds affected cell viability after 24h. However, at 10 µM there was a pronounced effect detectable in all three cell lines, especially for the negative control (Figure 14 B). Negative controls are often less thoroughly characterized and chemically optimized and can therefore results in more unspecific cellular responses. Both compounds had no impact on microtubule structures or cell membrane permeabilization (Figure 14 C). They represent a versatile set of tool compounds to investigate DDR1/2 and p38 function.
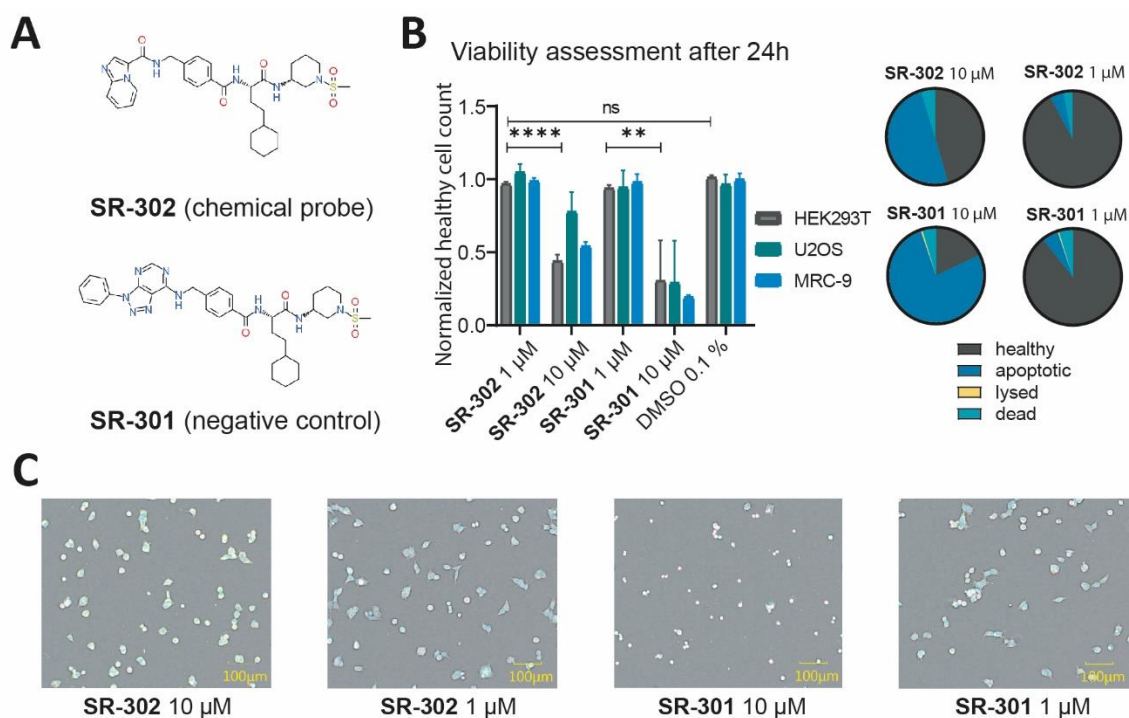
Figure 14: Live-cell high-content viability assessment of DDR/p38 probe compound SR-302 and corresponding negative control SR-301. This figure is adjusted from Röhm *et al.*[150]. **A.** Chemical structure of SR-302 and SR-301. Structures were used from the SGC DCP website (https://www.thesgc.org/donated-chemical-probes) **B.** Healthy cell count after 24 h of 10 µM and 1 µM compound exposure [SR-302, SR-301] normalized to healthy cells exposed to DMSO (0.1%) in HEK293T, U2OS, and MRC-9 cells. Error bars show standard error of the mean of quadruplicates. Significance was calculated using a two-way ANOVA analysis. Pie charts show fraction of healthy, apoptotic, lysed and dead cells after 24 h of 10 µM and 1 µM compound exposure [SR-302, SR-301] in HEK293T cells. **C.** Brightfield confocal image of stained (blue: DNA/nuclei, green: microtubule, red: mitochondria, and magenta: Annexin V apoptosis marker) HEK293T cells after 24 h of 10 µM and 1 µM compound exposure [SR-302, SR-301]).

## 4.4. Profiling compounds targeting dark kinase PCTAIRE on their impact on the cell cycle

Certain compounds exhibit their mode of action by effectively inhibiting their target protein within a specific phase of the cell cycle[108]. For example, some members of the cyclin-dependent kinases (CDKs) such as CDK2, CDK4 and CDK6, have been shown to play a pivotal role in cell cycle progression. Inhibition of these specific proteins can be used to block initiation and progression of specific cancer types, without harming healthy tissues[105, 153]. One subfamily of the CDKs is the PCTAIRE subfamily. So far, CDK16 is relatively understudied when compared to other kinases within this family[154]. Still, it was shown that CDK16 contributes to various

diseases such as breast, prostate and cervical cancer[155]. Thus, it is crucial to develop and characterize tool compounds and selective inhibitors targeting CDK16.

In the following publication, Amrhein *et al.*[156] (Appendix H) developed a series of 3-amino-1H-pyrazole-based kinase inhibitors, resulting in two highly potent compounds against CDK16 (42d and 43d), a member of the PCTAIRE subfamily. One of the compounds (43d) even exhibited exceptional selectivity against the PCTAIRE and the PFTAIRE subfamily against 100 other kinases.

"Discovery of 3-Amino-1H-pyrazole-Based Kinase Inhibitors to Illuminate the Understudied PCTAIRE Family"[156]

Until now, there is limited knowledge regarding the impact of the PCTAIRE family on the cell cycle. So far, it has been shown that the expression of CDK16 is highest during the S-phase of the cell cycle[157] and a knockdown in different cancer cells revealed a G2/M phase arrest followed by a p27 dependent apoptotic cell death[158]. Using the developed compounds, the influence of the PCTAIRE subfamily on the cell cycle using the fluorescent ubiquitination-based cell cycle indicator (FUCCI), was assessed. This system allows to distinguish between different cell cycle phases, using the licensing factor Cdt1 (chromatin Licensing and DNA Replication Factor 1) and its inhibitor Geminin[19]. Cells in S/G2/M will appear in green, cells in G1 in red and all cells in the transition state between G1 and S phase, show both GFP and RFP-labeled cells, so they will appear as yellow. As shown by Tjaden *et al.*[114], the FUCCI system can be combined with the viability assessment used in the *Multiplex* assay, which is based only on the nuclear morphology. I therefore additionally tested the influence of the compounds on cell viability. Compounds with high potency on CDK16 (21i, 42d, and 43d) (Figure 15 A) led to a concentration-dependent accumulation of cells in the G2/M phase over 24 hours (Figure 15 B). Compound (43d), exhibiting the highest selectivity for CDK16, produced the most impact in the cell cycle progression, which is consistent with the genetic knockout studies of Yanagi *et al.*[158] (Figure 15 C). The CDK2 inhibitor, Milciclib was used as a reference for G1 phase arrest[159]. They further explain a p57 dependent apoptotic cell death after a CDK16 knockout. There also was a decrease in cell viability found over time (Figure 15 D).
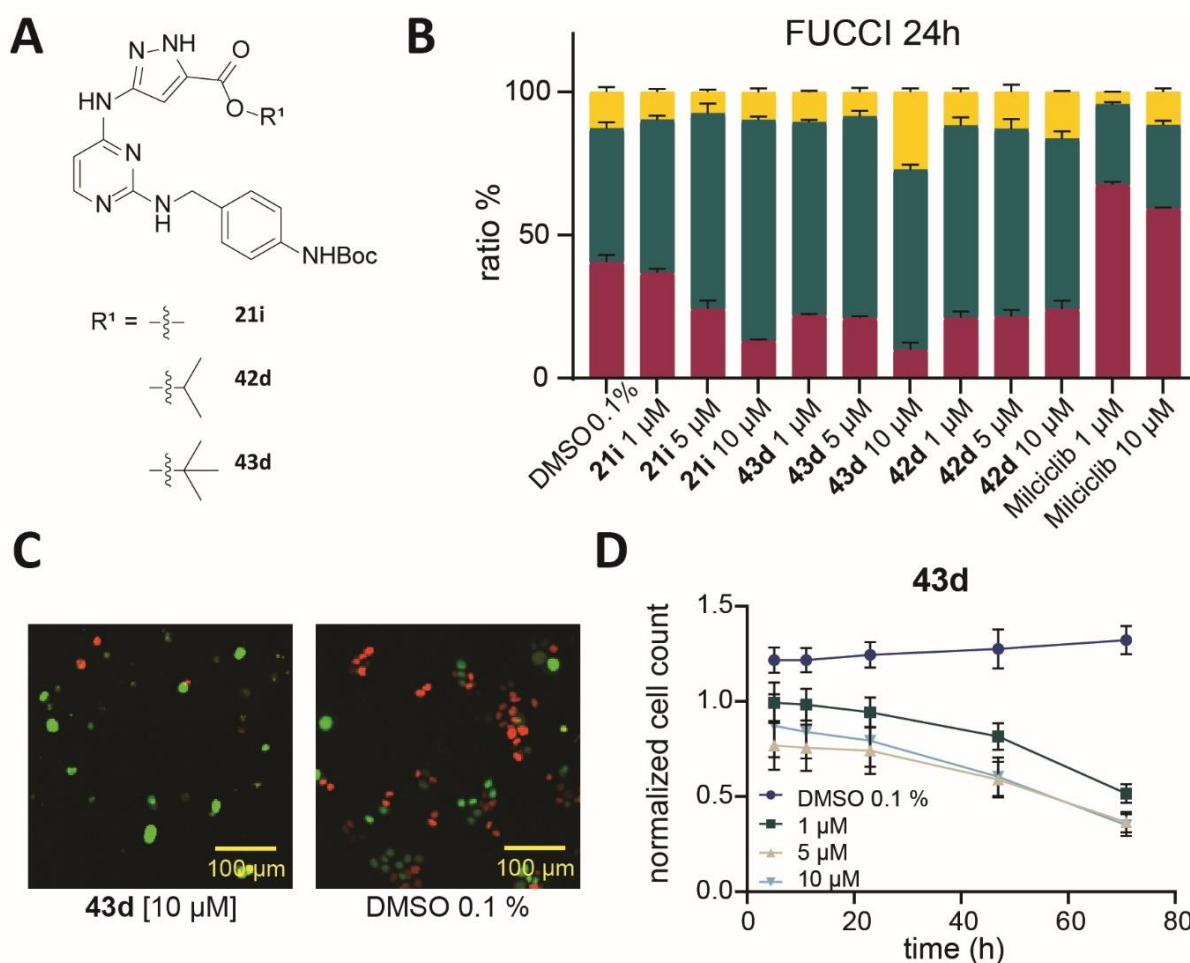
Figure 15: Cell cycle analysis and viability assessment using fluorescent ubiquitination-based cell cycle indicator (FUCCI) in HCT116 cells. This figure is adjusted from Amrhein *et al.*[156]. **A.** Chemical structure of *N*-(1*H*-pyrazole-3-yl)pyrimidine-4-amine-based inhibitors against CDK16 (21i, 42d and 43d) **B.** Fractions of red (G1), green (G2/M) or yellow (S) cells after 24 h of compound exposure (21i [1 µM, 5 µM, 10 µM], 43d [1 µM, 5 µM, 10 µM], 42d [1 µM, 5 µM, 10 µM], milciclib [1 µM, 10 µM]) in comparison to cells exposed to 0.1 % DMSO. Error bars show SEM of two biological replicates. **C.** Fluorescence image of HCT116-FUCCI cells after 48h of compounds exposure (43d [10 µM]) in comparison to cells exposed to 0.1 % DMSO. **D.** Normalized cell count of HCT116-FUCCI cells after 0h, 6h, 12h, 24h, 48h and 72h of compound exposure (43d [1 µM, 5 µM, 10 µM]) in comparison to cells exposed to 0.1 % DMSO. Error bars show SEM of two biological replicates.

4.5. Enhancing the detection of compounds inducing phospholipidosis through a computational approach and identifying chemical probes that trigger phsopholipidosis

A phenomenon impacting cell states is phospholipidosis (PL), which was recently brought into focus due to a perceived correlation of inhibitors to an antiviral activity against COVID-19[160]. Phospholipidosis is induced by an exposure of cells to external substances such as chemical substances. Due to an accumulation of phospholipids in acidic compartments such as lysosomes, the cell undergoes morphological changes and forms distinct structures such as lamellar bodies[21]. Given that compounds, capable of inducing PL, can interfere with diverse biological processes, developed drugs may lead to profound adverse effects including toxicity in liver and lungs. During drug development, drug candidates are therefore tested for their capability to induce PL[20, 161]. However, most current assays employed are endpoint assays and therefore do not assess this phenomenon in a kinetic manner. Therefore in Hu *et al.*[162] (Appendix I), I aimed to establish a live high-content cellular assay, which helped to develop a *machine learning* based algorithm to identify structural features critical for a PL induction and utilized this algorithm to validate various chemical probe compounds.

"A *machine learning* and live-cell imaging tool kit uncovers small molecules induced phospholipidosis"[162]

Several attempts have been undertaken to predict molecules that can induce phospholipidosis. The most commonly employed approach involves the analysis of the physicochemical parameters LogP and pKa[163]. Additionally, properties such as the net charge[164] or the volume of distribution (V_d) were considered[165]. To experimentally determine PL, electron microscopy (EM) has long been used as the gold standard[166]. Nevertheless, due to a low throughput, high costs and long timescales, the utilization of EM is limited. Consequently, diverse imaging techniques have recently gained popularity. Here, the most commonly used detection techniques are the use of fluorescently-labelled phospholipids such as (1-acyl-2-(12[(7-nitro-2-1-,3-benzoxadiazol-4-yl)amino]dodecanoyl) NBD-PE staining[167], nile red lipid stains[168] or the fluorescence microscopy-based lysosomal perturbation assay developed by Coleman *et al.* in 2010[169]. Unfortunately, a kinetic evaluation over an extended time period has not been adopted. In the presented publication (Appendix I), I developed a high-content assay to detect compounds that induce phospholipidosis in living cells. We optimized the LipidTox™ red

reagent[21, 170], where a PL induction enhances the fluorescent signal of the dye that is conjugated to phospholipids. I also incorporated this assay with the viability assessment of the *Multiplex* assay (see section 4.1. and 4.2.) to simultaneously validate the compounds impact on cell growth (Figure 16 A). To ensure an accumulation within the lysosomes, a LysoTracker™ probe was additionally added to the experimental setup. LysoTracker™ probes consist of a fluorophore linked to a weak base, which is trapped in acidic organelles[171] (Figure 16 B). The assay was validated using six known drugs that induce PL (Figure 16 C) and thresholds for compounds to be either categorized as "*inducers*" or "*non-inducers*" were defined. For U2OS cells, a compound was defined as inducer when showing a normalized PL intensity larger than 2, additional to a visual assessment of the images. Testing various other cell lines such as HEK293T and hepatoblastoma (HepG2) cells revealed, that under identical conditions the morphological differences are too substantial to establish a general analysis for all cell lines. Therefore, each cell line should be individually validated when it comes to PL.

To establish the *machine learning* algorithm, three datasets of different origin were used (Figure 16 D). The chemogenomic compound set comprised 213 diverse compounds targeting proteins of different families, including GPCR's (G-protein coupled receptors), kinases, epigenetic modulators and diverse others. The second set, introduced as the lysosomal modulation set, consisted of 99 compounds. While in this set less tool compounds were included, all of them are recognized to play a role in interfering with lysosomal processes. Both sets were tested for their ability to affect phospholipidosis in U2OS cells. Within the chemogenomic compound set 33 % and withinthe lysosomal modulation set 41 % of compounds were defined to be PL inducers. The third set resulted from a literature search, encompassing substances reported in nine articles representing PL inducing compounds. Following data curation, this third set was named curated data set. This set included in total 2308 compounds with 1683 being non-inducers and 625 being inducers. The curated data set was used to evaluate the effectiveness of a simple physicochemical property model (Ploemen's criteria using pKa and cLogP values)[163]. This resulted in 74 % of inducers and non-inducers that were correctly predicted. However, only 35 % of all PL inducers were identified, highlighting the limitation of this model.

To improve this prediction four distinct ML algorithms, namely k-nearest neighbor (KNN), support vector machine (SVM), random forest (RF), and eXtreme Gradient Boosting (XGBoost),

using MACCS and ECFP4 fingerprints as molecular representations were explored and evaluated on all three compound sets. The RF model using MACCS fingerprint exhibited superior performance over the others. As there is an ongoing debate about whether the cause of PL is related to the structural of substances, we investigated whether compounds with similar structures behave alike in their capability to induce PL. Our analysis suggests that even minor structural adjustments with minimal impact on physicochemical properties (pKa/cLogP) can lead to a different PL outcome. This was confirmed by testing chemical probe compounds with structurally close negative controls. This was highlighted by the three probes PPTN (an antagonist of the GPCR family member P2RY14), MS023 (targeting PRMT1, PRMT3, PRMT4, PRMT6, and PRMT8) and TP-472 (targeting BRD7 and BRD9) showed different PL outcomes than their negative control compounds (Figure 16 E).

These results demonstrated the need of investigating a time- and concentration-dependent phopholipidosis activity when it comes to compound annotation in cells.
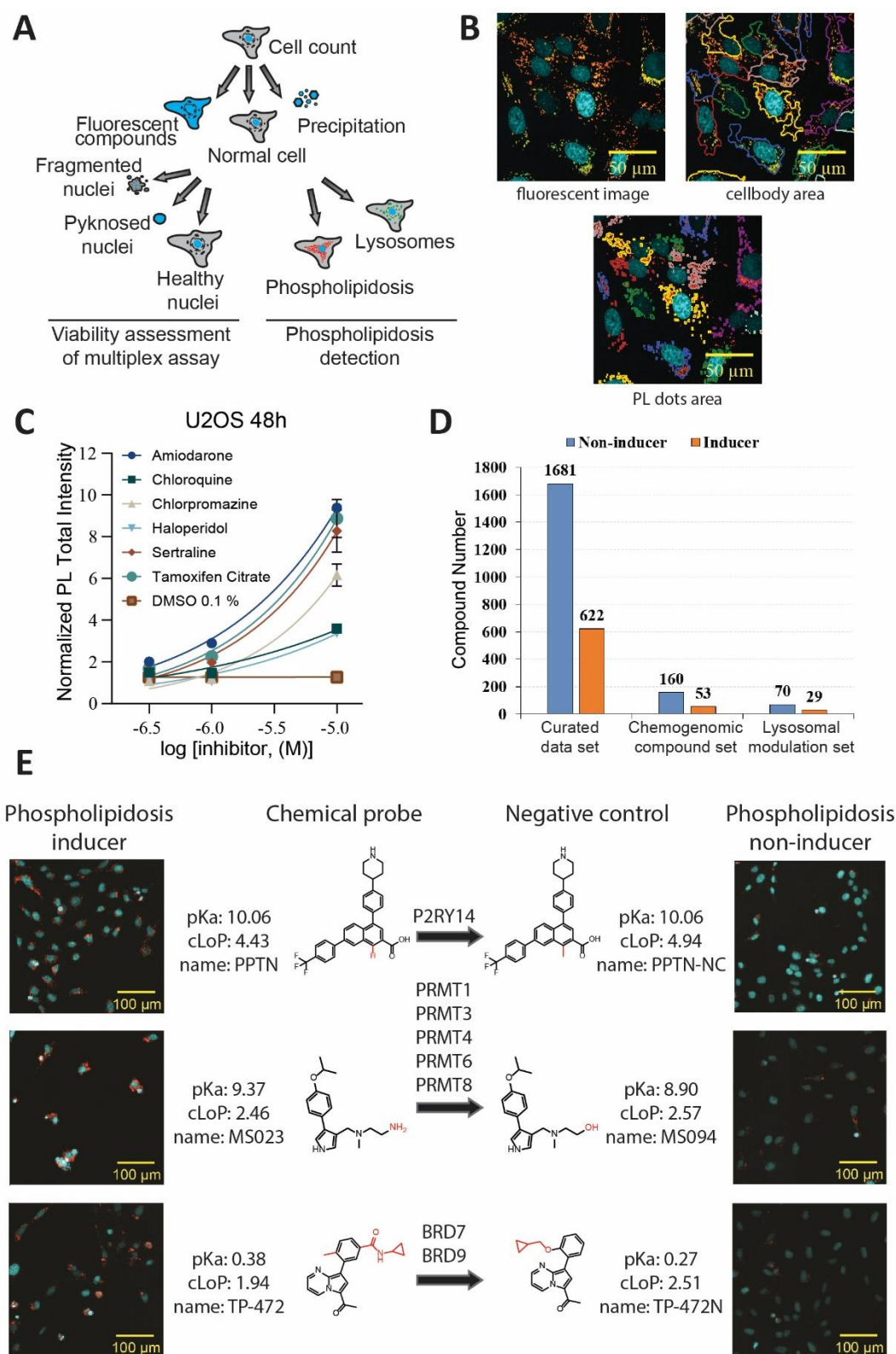
Figure 16: Phospholipidosis annotation in living cells to implement *machine learning* algorithm for PL detection. This figure is adjusted from Hu *et al*.[162] . **A.** Schematic representation of live-cell phospholipidosis protocol. **B.** Fluorescent images of stained (blue: DNA/Nuclei, green: lysosomes, red: phospholipidosis dots) U2OS cells and the detection of the cellbody area and phospholipidosis (PL) dots in Lysosomes after Cellpathfinder analysis. **C.** Normalized PL total intensity of U2OS cells exposed to 0.5 µM, 1 µM, and 10 µM of reference compounds

(amiodarone, chloroquine, chlorpromazine, sertraline, tamoxifen, and haloperidol) in comparison to cells exposed to 0.1 % DMSO after 48h. Error bars show standard error of mean (SEM) of four biological replicates. **D.** Bar diagram of different datasets used in this study (curated dataset, chemogenomic compound set, lysosomal modulation set). Bars with a blue color are showing the number of compounds defined as non-inducer, while orange bars show the number of PL-inducers of every compound set. **E.** Three analog pairs consisting of chemical probes and their corresponding negative controls are presented. Each pair is accompanied by the compound name and two calculated physicochemical properties (cLogP and pKa). The primary target gene names corresponding to each probe are provided in the center. Structural modifications between pairs are highlighted in red. Additionally, fluorescent stained image (blue: DNA/Nuclei, red: phospholipidosis dots) of U2OS cells for each compound is provided.

5. General Discussion and Outlook

Many small molecules used in disease models including cellular systems, are not fully characterized or validated, resulting in numerous misleading results and a large amount of false positive data[33]. This leads to data with poor reproducibility and considerable time and money being invested in follow-on projects that ultimately have no meaningful outcome. Unfortunately, this is affecting not only researchers in the field of chemical biology or the medicinal chemistry community, but also has strong impact on the whole drug development process[7]. A powerful tool to overcome this limitation, was the defining of guidelines for chemical tool compounds, with stringent criteria for selectivity and cellular potency[172]. Over the years, these criteria were continuously optimized by experts to enhance data generated with these compounds in diverse disease areas and to accommodate new compound modalities besides reversible inhibitors. In addition to assessing selectivity, potency and activity on the main target(s) of chemical tool compounds, off-target assessment plays a crucial role when it comes to compound characterization. An essential component of evaluating off-targets is the assessment of the impact of a compound on cell health. Especially, for oncological diseases where the main readout is based on impacting cell growth, it is crucial to distinguish if the resulting phenotype is connected to the main target or non-related cytotoxicity. For this purpose, easy-to perform and commercially available assays, that operate through the assessment of metabolic activity (MTT or alamarBlue®) or ATP production (CellTiter-Glo® assay) of the cells, are commonly used[61, 65, 66, 72]. They are widely used as they are easy to apply to different cells and are less expensive than more complex systems[173]. However, they have some disadvantages when it comes to cell health assessment. First, most of them are end-point methods, which is why they are often not suitable for the detection of small differences in proliferation rates. They don't have the ability to distinguish between cytostatic and cytotoxic effects of a compound, as cytostatic compounds show a more subtle growth inhibition, which may be missed using end-point methods[173, 174]. Second, they are frequently constrained in their sensitivity and as they rely on detection with a spectrophotometer or luminometer, there also is a risk of fluorescence interference by the compound to be tested, potentially leading to false results[173].

The most significant drawback, however, is that these assay systems cannot detect more complex cellular effects or distinguish between different cell death forms. They can only provide a straightforward assessment of the cell's metabolic activity, to make a quick statement about the impact of a compound on the overall cell viability. The impact of a compound on cell health is not trivial and can affect various cellular processes while also being influenced by them, which is why there is an urgent need to enhance the testing of cell viability in a more detailed manner. This can be obtained by using cell imaging techniques such as high-content imaging, to truly capture changes of the cells such as morphological changes, behavior of single cells or effects on cellular compartments[175].

The best strategy is to utilize a combination of different assays or to include diverse detection methods in one assay to comprehensively assess various aspect at the same time[173, 174]. The choice of assays should align with the specific purpose of the study, the conditions, or the cell types under investigation. Our purpose was to develop a strategy to test compounds in a high-throughput manner on their effects on cell health of different model cell lines to provide a basic assessment of these compounds on cell health and to enhance data quality of individual compounds and thereby reduce false positive or misleading data, aiming to improve the drug development process. I therefore established a pipeline consisting of a primary viability assessment followed by a high-content assay called *Multiplex* assay, with minimal impact on the exposed cells and a simple and fast protocol that can be adapted to detect diverse cellular features (see section 4.2.). In the following, I want to highlight the advantages and disadvantages of using the introduced pipeline, based on the EUbOPEN project, where we so far evaluated close to 1000 (925) compounds for their effects on cell health. Additionally, I want to elaborate on the reason, why most of the compounds resulted in a "*flag*" indicating, that they are not feasible to be used for target deconvolution alone or that specific phenotypic effects should be considered when testing them further (Figure 17 A). In the end I want to give an outlook on how the *Multiplex* assay can be further optimized and used for mechanistic studies.

The primary viability assessment included a label-free brightfield live-cell imaging approach, based on cell confluence detection (see section 4.2.). This assay was performed to exclude candidates that, compared to the negative control (DMSO), showed no influence on cell proliferation and are neither cytostatic nor cytotoxic. Conducting a primary screening

approach to reduce the amount of compounds that enter the second, more expensive and time intensive high-content screen, saves time and resources, as the probability that they exert unspecific effects are less likely. In this assay, however, only the effect on the general viability was tested, no phenotypic effects such as unspecific tubulin binding, which can also influence a follow up validation of a compound and force misleading compound data. To minimize the risk to miss out on compounds, that would be interesting to test in the second screen, we set the threshold relatively low, which resulted for the 925 compounds in over 65 % that were tested in the second screen, as they either resulted in a growth rate <0.5 or showed 'interesting' phenotypes after visual evaluation. The high ratio of compounds that were noticeable, might also be explained by the fact that many of them targeted kinases, such as Aurora kinase A (AURKA) which is important for mitotic cell proliferation[176] or DDR which is involved in the regulation of cell growth, differentiation and metabolism[177]. Kinase inhibitors are more likely to inhibit important signaling pathways resulting in cell death, in particular as many of these inhibitors have been developed as potential tools for oncological kinase targets. Additionally, the compounds were tested at a relatively high concentration of 10 µM. The purpose of this approach was to avoid categorizing the compounds as suitable for a compound set such as a chemogenomic library, despite the potential for inducing phenotypic changes at higher concentrations. Furthermore, in subsequent phenotypic studies, compounds are frequently employed at concentrations exceeding the recommended doses. Testing them at higher concentrations therefore enhances the reproducibility and comparability of different phenotypic data. All noticeable compounds were afterwards measured in the second screen, the *Multiplex* assay. High-content image-based methods such as the *Multiplex* assay, are excellent tools to discover different cell state changes. Each feature tested represents a compound characteristic that can influence diverse cellular reactions, can interfere with follow up validation of a compound and should be considered when using this compound in all kinds of applications. Out of these 600 tested compounds, 158 compounds were detected to influence different cell states with more than 50 % in more than one cell line. These so called "*flagged*" compounds represented 17 % of all compounds tested within the pipeline. 3 % of the compounds were flagged due to Hoechst High Intensity objects indicating a precipitation of the compounds or intrinsic fluorescence (Figure 17 B). Due to the reliance on fluorescence in most of the selectivity readouts, it is possible that the intrinsic fluorescence of compounds

would lead to falsely identified or unidentified targets. Another reason for Hoechst High Intensity objects is the precipitation of compounds in the cellular systems. This can result in diverse unspecific cellular reactions and also could lead to adverse events harming cells over time. One example of a compound showing these Hoechst High Intensity is Enzastaurin. Enzastaurin was initially identified as protein kinase C (PKC)-β selective inhibitor[178]. Later it was shown, that is also inhibits diverse other isoforms of protein kinase C (PKC)[178, 179], the glycogen synthase kinase 3 (GSK3) and ribosomal protein S6 phosphorylation, resulting in a lower vascular endothelial growth factor (VEGF) expression[180]. So far, it has been tested in diverse clinical trials with promising activity on a wide range on different cancers such as colon cancer and glioblastoma and showed pro-apoptotic growth-inhibitory effects[179-181]. In 2005, it was granted orphan designation for the treatment of glioblastoma. Nevertheless, there are multiple adverse events known such as high cytotoxicity and prolonged QT interval in guinea pig hearts and increased the action potential duration (APD) in a dose-dependent manner[182]. The initial failure to identify the main targets of the compound and even falsely identified targets of the compound, as well as the observed multiple side effect in different cell systems, may in part be attributed to the compound precipitation or intrinsic fluorescence. If the compound had been tested earlier for characteristics such as Hoechst High Intensity objects, it may have resulted in earlier abortion of the clinical studies, saving money, time and resources.

Interestingly, 6 % of the flagged compounds were negative controls. They are used as structural closely related controls to chemical tool compounds, with inactivity or less activity (100-fold less) on the main target. They are typically used to verify that an observed phenotype is driven by the target protein of the matching chemical probe[183]. To abrogate binding to the main target, often just small chemical modifications are engineered into the negative controls. The change of the structure by a single atom can already change the activity against a target up to 80 %[183]. This occurrence is not limited to the main target but also extends to the off-target profile of a compound. As negative controls are typically used to confirm a phenotype induced by a chemical tool compound, rather than to explore its own target space, they are often less characterized than their matching probe compound in terms of selectivity and potency. This can influence the resulting phenotype and can lead to wrong gene-phenotype associations. It is therefore not surprising that negative controls exhibit more non-specific effects on cells. For

example RL2578, the negative control for the AKT1/2 inhibitor Borussertib, affected microtubule integrity at 1 µM[184, 185] or BAY-294, the negative control of the SOS1 inhibitor BAY-293[186] precipitated after 24h in HEK293T cells.
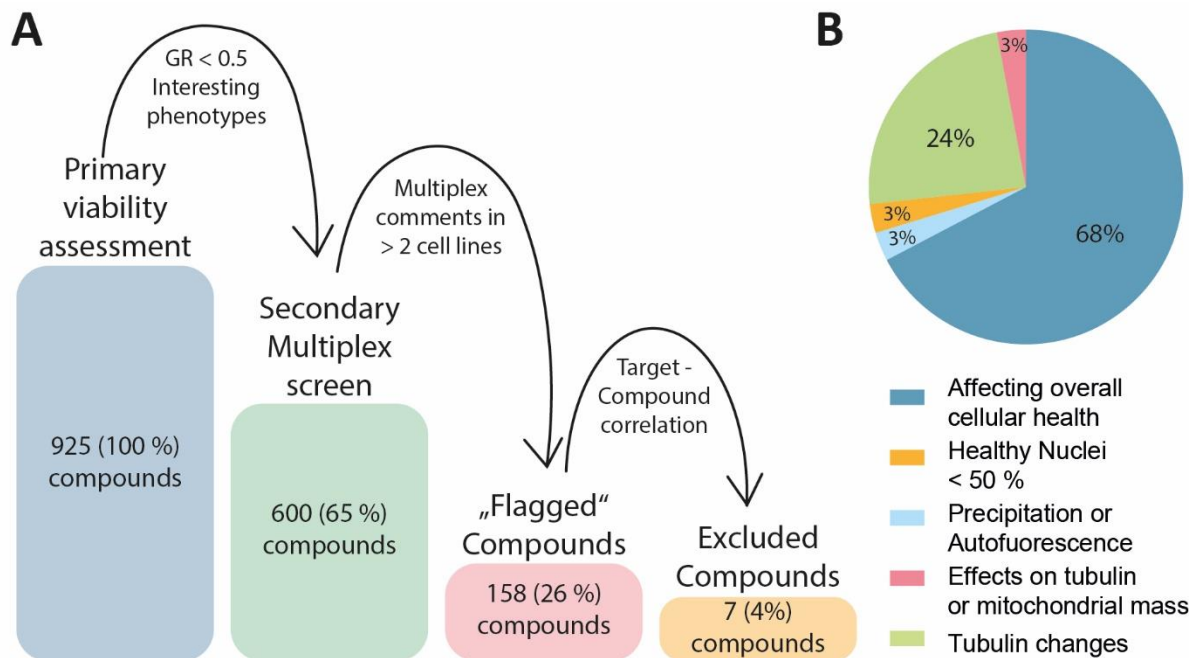
Figure 17: Viability assessment pipeline resulting in 158 "*flagged*" compounds. **A.** This figure shows the workflow of the viability assessment pipeline used to characterize cellular quality of diverse compounds. Out of 925 compounds tested in the primary viability assessment, 600 compounds were "noticeable" with a GR < 0.5 or interesting phenotypic appearance. They were tested in the *Multiplex* assay and "*flagged*" when they influenced different cell states after 24h to more than 50 % in more than one cell line. **B.** Distribution of features for a compound to be "*flagged*" after *Multiplex* assay evaluation. This figure was created using Adobe Illustrator v26.5.

As it was already shown, that just by counting of the cell nuclei alone is the most accurate method to detect cell viability[174], we implemented this in our assay setup. We were able to demonstrate that a gating of the cells based on their nuclear morphology results in similar IC$_{50}$ values compared to the gating based on the whole cell morphology. For maximum efficiency in utilizing one channel, we further optimized the gating protocol for the cell nuclei. Next to the detection of precipitation or intrinsic fluorescence and the influence on cell viability, we can make predictions about the cell death mechanism leading to the viability decrease and even detect mitotic cell cycle arrest. However, as one of the main readouts of the assay is based on one distinct channel, it is important to use the right concentration of Hoechst33342, as it

was shown to be toxic at concentrations higher 1 µM[187, 188]. Additionally, compounds that interfere with this channel, showing intrinsic fluorescence at the same wavelength or destroying cellular DNA can be much harder to evaluate as they can result in false positive results. A major advantage of using just one distinct channel for multiple cellular investigation was that it, enabled us to simultaneously assess additional cell health parameters in the same assay setup. For the *Multiplex* assay we decided to implement the detection of cytoskeletal changes, increase of the mitochondrial mass, membrane permeabilization and apoptotic cell death.

Assessing diverse features at the same time, enabled us to observe, that only 3 % of all compounds tested in the *Multiplex* assay, affected the viability without influencing the other examined phenotypic effects such as tubulin structures or mitochondrial damages. Viability and phenotypic effects such as affecting tubulin structures or mitochondrial health are influenced by each other. If a compound affects the overall cell viability or if different cell death mechanisms such as apoptosis or necrosis are triggered, diverse cell organelles are naturally affected and can undergo structural changes. The majority of the flagged compounds therefore affected the overall cellular health (68 %), which included different phenotypic effects such as mitochondrial changes and tubulin interference. Additionally, a strong dependence between phenotypic effects and cell viability was also shown as none of the compounds were flagged due to an influence on mitochondria without decreasing the overall cell viability. Mitochondria are essential for various cell functions, consequently most of the compounds that affect mitochondrial health will also impact the overall cell health.

To distinguish if a phenotypic effect results from an effect on cell viability or viability is influenced due to an effect on cell organelles, such as microtubules or mitochondria, compounds could be tested in a dose-dependent manner over time. To illustrate this dependency, I want to introduce a project initiated by Martin Schwalm, where I tested, among other things, potential proteasome inhibitors in HEK293T cells over 24h in a concentration-dependent manner, using the *Multiplex* assay (Figure 18). In the *Multiplex* assay, I tried to distinguish between phenotypic effects and cell viability by including the tree-based principle, where cells were grouped as "healthy" before being gated in different phenotypic groups. This approach allowed me for example to detect changes in tubulin structures that are not undergoing cell death, as identified by a pyknosed or a fragmented nucleus. For example,

Bortezomib, a proteasome inhibitor that was approved by the FDA for the treatment of multiple myeloma and mantle cell lymphoma[189, 190], reversible inhibits the chymotrypsin-like activity of the 20S proteasome[191, 192]. It was shown that Bortezomib induces endoplasmic reticulum (ER) stress and leads to cytochrome c release and activation of effector caspases, resulting in apoptotic cell death[191]. Another compound, RA190 has originally been published to bind cysteine 88 of the ubiquitin receptor RPN13 in the 26S proteasome[193], but later found to rather be a very unspecific chemical compound targeting diverse proteins[194]. We compared the concentration and time-dependent effect of both compounds on the cell count against two phenotypic properties, including the influence on mitochondrial mass and tubulin integrity. For Bortezomib, I observed an impact on both mitochondria and microtubules at concentrations, at which overall cell viability was comprised as well (Figure 18 A). This highlights that there is a mutual influence between the effect on cell viability and phenotypic changes of the cells and that it is often difficult to distinguish which of these characteristics is the primary cause of the observed phenotype. In case of the second tested compound, RA190, I observed that alterations of microtubule and mitochondria occurred at earlier time points and lower concentrations than impacting cell viability (Figure 18 B). At 1 µM RA190 influenced tubulin integrity and impacted the mitochondrial mass already after one hour. At the same concentration, however, a decrease of the cell count on the same extend was only observed after 6 hours.
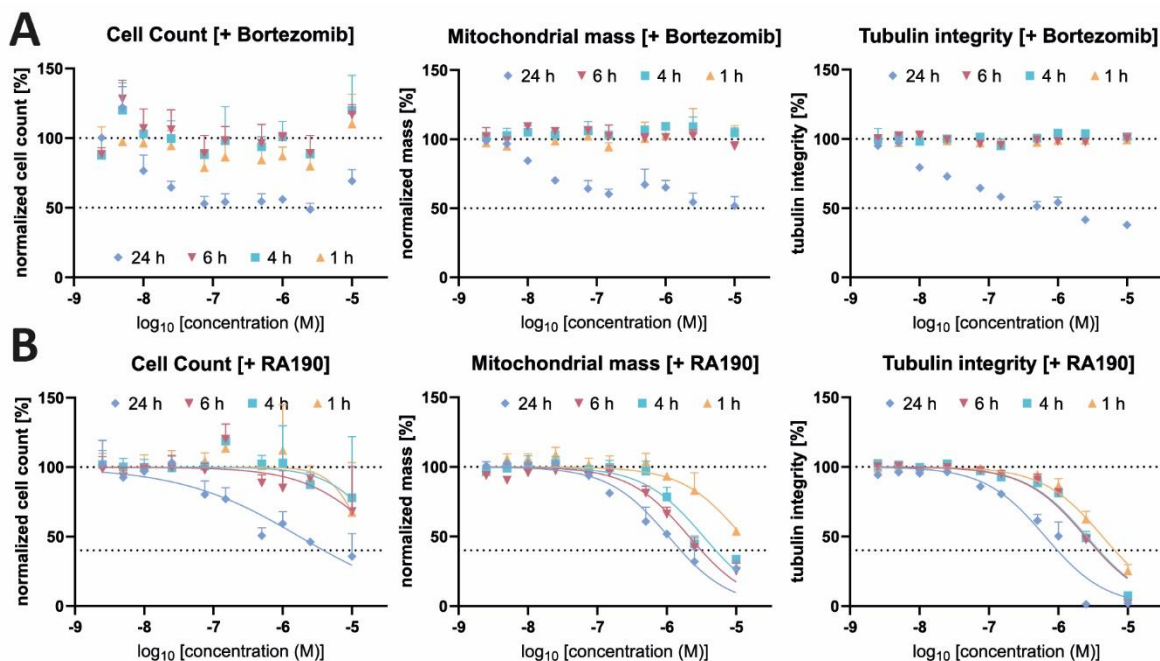
Figure 18: Evaluating time and concentration dependent influences on cell states. This figure was adjusted from Schwalm *et al*. (to be published soon). **A.** Normalized cell count, effect on mitochondrial mass and tubulin integrity after 1h, 4h, 6h and 24h of HEK293T cells exposed to Bortezomib (25 nM – 10 µM). Data was normalized against cells exposed to 0.1 % DMSO. Error bars show standard error of mean (SEM) of biological duplicates. The overall cell count and the phenotypic properties are comparably affected at the same concentrations and time points. **B.** Normalized cell count, effect on mitochondrial mass and tubulin integrity after 1h, 4h, 6h and 24h of HEK293T cells exposed to RA190 (25 nM – 10 µM). Data was normalized against cells exposed to 0.1 % DMSO. Error bars show standard error of mean (SEM) of biological duplicates. The phenotypic properties are influenced at earlier time points and lower concentrations than the overall cell count. This figure was created using Graph Pad Prism v8.4.3.

Using end-point methods or testing the compounds only after a specific time-points at single concentrations, we could not be able to distinguish if the phenotypic effect was the cause of harming the cells or the reason for it. The *Multiplex* assay serves as a versatile tool to distinguish these small differences in the same cell after different time points and include the impact on cell viability when it comes to phenotypic characterization. We were therefore able to see that almost a quarter of the compounds (24 %) affected tubulin structures before they affected the overall cell viability. It can therefore be concluded that for these compounds cell death was a consequence of the compound's effect on tubulin. As tubulin binding can also be a desired effect of a compound, especially in cancer therapy to disrupt cellular functions, for some of the compound the observed effect was already known. For example, CGP-60474, a potent inhibitor of cyclin-dependent kinases and protein kinase C (PKC), had an impact on tubulin structure higher 50 % in more than one cell line and was therefore "*flagged*" after *Multiplex* assay assessment[195, 196]. As CDKs have been described to influence diverse cellular processes such as regulation of the cell cycle, splicing, transcription or DNA repair, this effect is due to the inhibition of the targets. However, for some compounds the impact on microtubule was so far not known, such as for the chemical probe BAY-784. BAY-784 selectively targets the gonadotropin releasing hormone receptor (GNRHR)[197, 198]. Maybe the observed effect was a result of affecting tubulin within a signaling complex known as the signalosome, consisting of GNRHR, diverse kinases and tubulin[199]. However, for some compounds the impact on microtubules was not known, such as for the chemical probe BAY-784. BAY-784 selectively targeted the gonadotropin releasing hormone receptor (GNRHR)[197, 198]. Maybe the observed effect was a result of affecting tubulin within a signaling complex known as the signalosome, consisting of GNRHR, diverse kinases and tubulin[199]. However, it could also stem from non-

specific binding to a segment of microtubules. The assay relies on an implemented *machine learning* algorithm, that detects changes of tubulin in comparison to "*normal*" looking cells. Consequently, it can detect that the tubulin structure is affected but at the chosen magnification for the assay a precise determination of how the tubulin structure is influenced was difficult. Changes of microtubule structures can manifest itself in diverse ways. Compounds can either inhibit the disassembly or the assembly or bind to different parts of microtubules, as there is a complex interplay between various molecules[200, 201]. Compounds without a known indication regarding their effect on microtubules, whether related to the target or known off-targets, were therefore not directly be excluded from the chemogenomic set. Instead, these findings were shared with the community via BioImage Archive (https://www.ebi.ac.uk/biostudies/bioimages/studies/) and the Gateway (https://gateway.eubopen.org/) providing valuable information for other researchers to compare their data when observing similar effects. For a more in-depth understanding of a compound's impact on microtubule, higher magnifications or alternative biochemical methods such as electron microscopy prove to be a more efficient tool[202]. Hence, a next step would be to further investigate these compounds to offer a more comprehensive assessment about their precise influence on microtubules and the underlying functionality.

Live cell imaging allows the detection of kinetic changes and better comparability of compound characterization. So far, most of the well-established high-content assays, which examine cellular changes from a multiplex perspective, are based on fixed samples, detecting only one distinct time point. One prominent example is the cell painting assay, which was introduced in 2013[203]. Here, eight components (DNA, cytoplasmic RNA, nucleoli, actin, Golgi apparatus, plasma membrane, endoplasmic reticulum and mitochondria) are detected utilizing six cellular stains, imaged on five fluorescent channels. It is a high-throughput method to capture various cellular features, which are then used to generate morphological profiles of each compound[204-206]. In contrast to the *Multiplex* assay, this assay provides many more cellular features and multiparametric information. However, using fixed samples limits the ability to capture dynamic, real-time cellular processes and do not encompass compound effects at different time-points.

Next to the time-dependence another crucial aspect of compound characterization and to distinguish a phenotypic effect from an unspecific effect, is the use of the correct

concentration, particularly when testing chemical tool compounds. Chemical tool compounds selectively target one or more proteins at one specific concentration range. They are used for mechanistic and phenotypical studies to interpret the biology of their target(s). To decipher a target-phenotype relationship or use them in general for mechanistic studies, it is therefore crucial to use them at their recommended concentration[5, 10, 44]. When used at the wrong concentration, even selective compound will interact with different targets or result in diverse unspecific or toxic effects[5]. Therefore, chemogenomic compounds are annotated at their recommended concentration and additionally at a 10-times higher than the recommended concentration[16]. For the chemical quality pipeline, all compounds are tested at the same two concentrations for better comparison, easier execution and robustness. However, to evaluate if the resulting effects are on-target or off-target related, one should always consider the target affinity and therefore test the compound rather at their recommended concentration. For example, if the concentration of a compound against its main target is in the low nanomolar range, the cellular annotation of this compound should also be performed at this concentration range. This was especially visible, when evaluating chemical probes and their matching negative control on phospholipidosis induction (section 4.5.). Six compounds out of 31, showed a phospholipidosis induction at their recommended concentration[162], which is an unspecific side effect that can lead to a variety of phenotypic changes or clinical adverse events[125]. A study of Sterling *et al.*[5] in 2023 also revealed that chemical probes are frequently used incorrectly. They examined 662 publications in which chemical probes were used. Surprisingly, 22 % employed these chemical probes at concentrations higher than recommended and 58 % neglected to include an inactive control or an orthogonal compound in addition to the probe compound. Therefore, such studies cannot be used to determine whether the resulting effect, such as a specific phenotype, is due to the primary target or some off-target.

The effect of a compound on cell health is an important parameter when it comes to compound annotation in cells. Nevertheless, other unspecific compound characteristics such as the influence on the cell cycle, phospholipidosis or more complex cell death mechanisms such as autophagy can influence compound activities as well. Depending on the research question asked, the purpose of the assay and the available characterization data of a compound, the specific cell system relevant for the disease or biological functional of the compound should be tested. The *Multiplex* assay protocol is a useful tool in these different

scenarios as it is easy adaptable to include other fluorescent biological reporters or to test diverse cell systems. By assessing a range of cell lines, more potential effects and cell-type specific responses can be detected and compared. In the *Multiplex* assay standard setup, therefore three different cell lines (HEK293T, U2OS and untransformed MRC-9 fibroblasts) are tested. This was decided, as they are often used as a standard in the field of medicinal chemistry, enhancing the comparability of different assays[207]. Additionally, these cell lines are often already used in diverse laboratories, which therefore enables an easy set-up of the assay in a different laboratory. As each cell lines possesses unique genetic properties they can respond to external stimuli, such as chemical compounds, in different manners[208]. Testing these three cell lines, already allowed us to observe variations in the cellular responses to a given compound. This was not surprising, as the cell lines naturally differ in their morphology and proliferation rate. This is also important for the image-based evaluation of the cells, as the algorithm incorporates morphological features when gating the cells into different groups. Therefore, every cell line needs to have their own, individual image analysis, that was specifically trained using cell images of the distinct cell line. I could also observe that cell lines having a similar proliferation rate are more likely to cluster together, which was demonstrated as the cancer-based cells HEK293T and U2OS were more likely to induce similar characteristics in comparison to the untransformed fibroblasts. Ideally, one would select for every compound (group of compounds) a cell system that is the most relevant for the target space, such as a potential disease model and compare this data to standard cell lines. If the protocol is used to investigate other features or cell lines of interest it must be adapted accordingly. An important aspect is that the dyes should not interfere with the cellular responses and viability itself. Especially in applications with living cells, it is crucial to determine the optimal concentration of the biological reporter used, such as different fluorescent dyes, that ensures a robust signal-to-noise ratio with minimal impact on the cells[134]. This means, when the protocol is adapted to other dyes or other cell lines, the impact of the biological reporters on the new system has to be evaluated.

As an intriguing direction for the future, it would also be fascinating to explore whether the setup of the *Multiplex* assay could be adapted to more complex systems such as patient derived cell systems or three-dimensional (3D) cell cultures. In comparison to two-dimensional (2D) cell cultures, which have many limitations and cannot represent a physiological response,

more complex systems such as spheroids or organoids, are used to better mimic *in vivo* conditions[209, 210]. Although, with these complex systems, cell-cell interactions, tumor characteristics, metabolic processes and other cell environmental factors can be investigated, they are still more time consuming, require more complex assay setups and often have limited throughput, especially when it comes to cell imaging[209-211].

In various projects, I could demonstrate the significance of compound annotation in a cellular context and how high-content imaging serves as a valuable tool for examining these diverse effects of compounds on cell states. With the growing importance of high-throughput screening methods and computational approaches, more complex data can be generated in a faster and easier way. However, as these methods are based on automated analysis protocols with high demands on computational power, such assays are often associated with high costs and expensive and high-maintenance equipment. So far the images are evaluated using the CellPathfinder software (Yogogawa)[212]. This software has a wide range of functions, such as *machine learning* detection and in newer versions even *deep learning* functions. However, the software is not publicly accessible and does not have the most intuitive user interface. To make such an assay and the analysis wider accessible, it would be desirable to implement the analysis using open-source software such as Image J (https://imagej.net/) or Cell profiler (https://cellprofiler.org/). To enable people to utilize the generated data in an open science context, we have uploaded all data via freely accessible platforms such as ChEMBL (https://www.ebi.ac.uk/chembl/) and the BioImage Archive (https://www.ebi.ac.uk/bioimage-archive/). This allows users to obtain necessary information about a compound of interest and utilize the data for their own analyses. Data availability is also important to enhance the data landscape of the compounds. It can be utilized not only by scientists in the field of biology and chemistry but also by computational scientists. The recent emergence of artificial intelligence (AI)-based data analysis plays a significant role in accelerating and simplifying the drug development process. Often, computational scientists can use the already available data to evaluate diverse features and enhance compound knowledge. Nevertheless, these approaches require high computational power and extensive knowledge of the available data and usage and it is often hard to make these algorithms applicable and comparable. The data used for such AI-based analysis need to be highly reproducible and robust. Therefore, it would be desirable to use standardized equipment and standardized analysis protocols, which is often

challenging when especially for example *machine learning* algorithms that were trained using different example images. For future directions a combination of computational approaches together with experimental data would be desirable. For this, computational experts and researches should combine forces to improve the drug development process.

To use small molecules, such as chemical tool compounds for the development of a new drug or to decipher biological signaling pathways, it is crucial to annotate them regarding their selectivity for different proteins and their potency on the main target, and on their diverse cellular properties and off-target profiles. For this, it is crucial to test a compound in diverse cell systems over different time points and at a meaningful concentration range. The here introduced pipeline, serves as a versatile tool, to annotate diverse cellular compound characteristics using an easy and fast protocol that can be adapted to diverse research questions and cell systems. It enhances the characterization process of compounds, as it can be used in a middle-throughput manner, testing hundreds of compounds at the same time. In order to promote the use of this pipeline within the scientific community, as a fundamental method for evaluating the effect of small molecules on their cellular effect, I have made all protocols and analysis setups utilized available and uploaded all data regenerated so far, for people all over the world to be used.

6. Appendix

Appendix A: Image-Based Annotation of Chemogenomic Libraries for Phenotypic Screening

Contribution Menge, A.: Conceptualization, methodology and formal analysis, validation, data curation, writing—original draft preparation, writing, review and editing.

molecules

MDPI

Article

Image-Based Annotation of Chemogenomic Libraries for Phenotypic Screening

Amelie Tjaden [1,2], Apirat Chaikuad [1,2], Eric Kowarz [3], Rolf Marschalek [3], Stefan Knapp [1,2], Martin Schröder [1,2,*] and Susanne Müller [1,2,*]

1. Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Max-von-Laue-Str.9, 60438 Frankfurt, Germany; tjaden@pharmchem.uni-frankfurt.de (A.T.); chaikuad@pharmchem.uni-frankfurt.de (A.C.); knapp@pharmchem.uni-frankfurt.de (S.K.)
2. Structural Genomics Consortium, BMLS, Goethe University Frankfurt, Max-von-Laue-Str. 15, 60438 Frankfurt, Germany
3. Institute of Pharmaceutical Biology, Goethe University, Max-von-Laue-Str.9, 60438 Frankfurt, Germany; kowarz@em.uni-frankfurt.de (E.K.); rolf.marschalek@em.uni-frankfurt.de (R.M.)
* Correspondence: m.schroeder@pharmchem.uni-frankfurt.de (M.S.); susanne.mueller-knapp@bmls.de (S.M.)

Abstract: Phenotypical screening is a widely used approach in drug discovery for the identification of small molecules with cellular activities. However, functional annotation of identified hits often poses a challenge. The development of small molecules with narrow or exclusive target selectivity such as chemical probes and chemogenomic (CG) libraries, greatly diminishes this challenge, but non-specific effects caused by compound toxicity or interference with basic cellular functions still pose a problem to associate phenotypic readouts with molecular targets. Hence, each compound should ideally be comprehensively characterized regarding its effects on general cell functions. Here, we report an optimized live-cell multiplexed assay that classifies cells based on nuclear morphology, presenting an excellent indicator for cellular responses such as early apoptosis and necrosis. This basic readout in combination with the detection of other general cell damaging activities of small molecules such as changes in cytoskeletal morphology, cell cycle and mitochondrial health provides a comprehensive time-dependent characterization of the effect of small molecules on cellular health in a single experiment. The developed high-content assay offers multi-dimensional comprehensive characterization that can be used to delineate generic effects regarding cell functions and cell viability, allowing an assessment of compound suitability for subsequent detailed phenotypic and mechanistic studies.

Keywords: phenotypic screening; high content imaging; chemogenomics; machine learning; cell cycle

1. Introduction

Phenotypic screening has recently experienced a resurgence in drug discovery after many years of focus on target based approaches [1]. In particular, methods such as cell painting [2–4] or phenomics are gaining interest due to their ability to detect disease relevant morphological and expression signatures. These exciting new technologies provide insights into the biological effects of small molecules on cellular systems and the suitability of identified hits for translational studies. One of the main advantages of phenotypic screening lies in the potential of identifying functionally active chemical modulators without the need to know their precise mode of action (MoA). However, the lack of detailed mechanistic insight complicates the rational development of identified hit matter and validation studies [5]. One way to circumvent these complications is the use of better annotated chemical libraries, consisting of highly target-specific chemical probes [6–8] or chemogenomics libraries which contain well-characterized inhibitors with narrow but not exclusive target selectivity [9,10]. In particular, the latter have gained increasing interest as a new approach in drug discovery [11,12] as chemogenomic libraries

may cover a large diversity of targets and a larger fraction of druggable proteins. Thus, chemogenomic compounds (CGCs) can supplement chemical probe collections that are not available for many targets due to their stringent quality criteria [13]. In cellular studies, the use of several CGCs directed towards one target but with diverse additional activities, will allow deconvolution of phenotypic readouts and identification of the target causing the cellular effect [14,15]. In addition, compounds from diverse chemical scaffolds may enable an easier identification of off-targets from different families. Further validation such as proteomic-based approaches or quantitative structure-activity relationships (QSAR) may be required [16]. The importance of chemogenomics for drug development has recently been demonstrated by a call of the Innovative Medicines Initiative (IMI), which resulted in the funding of the EUbOPEN project. One aim of this project is to assemble an open access chemogenomic library covering more than 1000 proteins by well annotated CGCs as well as chemical probes [17]. The expansion of this CGC collection to cover the entire druggable proteome will be the goal of the Target 2035 [18].

While target selectivity is an important parameter, there is a need for a comprehensive annotation of CGCs in terms of quality of the used chemical matter, such as structural identity, purity and solubility. In addition, the effects of CGCs on basic cellular functions such as cell viability, mitochondrial health, membrane integrity, cell cycle and interference with cytoskeletal functions which may be affected by non-specific binding of CGCs to tubulin should be considered [19]. Computational as well as screening approaches have been employed to predict the (unspecific) toxicity of libraries used for screening [20]. Although it is not always easy to distinguish between on-target and off-target effects in a cell viability assay, adding information on chemical and biological quality to CGC libraries will help to differentiate between target specific and unspecific effects [10]. New technology developments such as automated image analysis systems and machine learning algorithms enabled high-content techniques to become the method of choice for the essential annotation of chemogenomic libraries. Here, we present a modular live-cell high-content cellular viability assay, which we expanded to include assessment of CGC effects on the cell cycle, tubulin, mitochondrial health and membrane integrity. In contrast to the Cell Painting assay, which captures a multitude of morphological features of fixed cells at a given time point and requires extensive downstream analysis [3], the purpose of our assay is to describe cell health in living cells, providing the opportunity for real-time measurement over a long time-period. The modular nature of the assay offers the opportunity for an expansion such as adding compound-safety assays and other cellular stress reporter systems without the need for additional informatics capacities [21].

2. Results

2.1. Optimization of HighVia Protocol and Validation of Cell Staining Dyes

Analyzing cytotoxicity at multiple time points improves the annotation of small molecules and facilitates distinguishing between primary and secondary target effects. In this report, we improved our previously published, single time point protocol [22] to provide a more continuous readout. Live-cell imaging using fluorescent dyes for an extended period of time requires low concentrations of dyes that do not interfere with cellular functions yet provide fluorescent signals that are sufficiently high for robust detection. Therefore, we first optimized the concentration of the DNA-staining dye Hoechst33342 and determined 50 nM as the minimal concentration that still yielded robust detection of nuclei in HeLa cells (Supplementary Figure S1). Previous studies have identified the toxicity level of nuclear stains such as Hoechst33342 at concentrations around 1 µM [23,24]. We tested in a cell viability experiment using U2OS cells and the alamarBlue™ dye (alamarBlue HS reagent, ThermoFisher, Massachusetts, United States) whether Hoechst33342 at concentrations below 170 nM resulted in reduced viability (Figure 1A/Supplementary Table S1). Additionally, we assessed in this experiment the potential effects on cell viability of other live-cell dyes such as the mitochondrial stain MitotrackerRed® and the taxol-derived tubulin dye BioTracker™ 488 Green Microtubule Cytoskeleton Dye. Gratifyingly, none of the dyes

exerted any significant impairment of cell viability at the proposed assay concentration over a time period of 72 h (Figure 1A). To exclude the potential influence of multiple dyes at their given concentration would influence viability, we tested different combinations of dyes in U2OS cells using an orthogonal high content readout (Figure 1B/Supplementary Table S1). Consistent with the data of the single dye experiments, none of the dyes or their combination inhibited cell viability.

Encouraged by these results, we assessed whether by lowering the dye concentrations of the previously published HighVia protocol (workflow of HighVia protocol see Figure 1C), this method could be adapted to allow a continuous readout (workflow continuous protocol see Figure 1D). In addition to the dyes used in the HighVia protocol, we included MitotrackerDeepRed® to measure the mitochondrial content and thus named the protocol 'HighVia Extend'. Changes of mitochondrial mass are indicative of certain cytotoxic events such as apoptosis [25,26]. Cells were detected as previously described [22] and gated into five different populations using a supervised machine-learning algorithm (Figure 1E/Supplementary Table S7. We chose nine reference compounds as a training set (Supplementary Table S2) for the assay setup, which was tested in three different human cell lines: human embryonic kidney cells (HEK293T), osteosarcoma cells (U2OS) and non-transformed human fibroblasts (MRC9). These reference compounds covered multiple modes of actions including topoisomerase inhibitor camptothecin, which triggers apoptotic cell death by inducing strand breaks in chromosomal DNA [27], the BET bromodomain inhibitor JQ1 [28], the mTOR inhibitor torin [29] and the glycosidic drug digitonin, a detergent used to permeabilize cell membranes [30] (Figure 1F). We found that the new continuous assay format captured the kinetics of the selected diverse cell death mechanisms: the cell-membrane permeabilizing agent digitonin as well as the multi-kinase inhibitor staurosporine and the ATM/ATR inhibitor berzosertib displayed rapid induction of cytotoxicity, while inhibitors of epigenetic targets JQ1 and ricolinostat showed slower and less pronounced cytotoxic effects (Figure 1F), consistent with previous reports [31,32]. Treatment with the non-selective CDK inhibitor milciclib, the mTOR inhibitor torin and the tubulin-disassembly inhibitor paclitaxel resulted in cytotoxic response with intermediate kinetics. IC_{50} values of the different compounds over time are compiled in Supplementary Table S2.

Consistent with the overall cell count of healthy cells, the population gating also followed different kinetic profiles, exemplified by camptothecin (Figure 1G). These data suggested that the continuous format of the HighVia Extend facilitated the assessment of time-dependent cytotoxic effects of small molecule compounds.

2.2. Investigation of Nuclear Properties

While analyzing data from the continuous experiment, we identified a strong correlation between the overall cellular phenotype (categories: "healthy", "early/late apoptotic", "necrotic", "lysed") and the nuclear phenotype defined as either "healthy", "pyknosed" or "fragmented" (Figure 2A). To further test whether the gating based on the nuclear phenotype alone resulted in similar cytotoxicity profiles, we compared the calculated IC_{50} values of the aforementioned nine reference compounds gated either as described above (based on entire cellular phenotype) or based on the nuclear phenotype alone (features used for machine learning algorithm described in Supplementary Table S7). We found that the time-dependent IC_{50} values and the maximal reduction in the healthy cell population were highly comparable between these gating methods (Figure 2B,C/Supplementary Figure S2A–F). Additionally, the overall population distribution profiles from both gating protocols were highly similar (Figure 2D,E). The dependence on only one fluorescent channel might however increase the risk of assay interference of compounds with similar fluorescent properties such as berzosertib or with poorly soluble small molecules such as itraconazole that exhibit high fluorescent background (Figure 2F, Supplementary Figure S3). In order to minimize the risk of such interferences, we included an additional layer of gating to the protocol. In the first step, all fluorescent objects in the corresponding channel were classified either as

"nuclei", independent of their phenotype or as "high intensity objects" that detected both fluorescent compounds as well as precipitations (Figure 2G,H). We noticed that a limitation of this method was that we were not able to readily distinguish between pyknosed nuclei of mitotic cells or condensed nuclei of apoptotic cells. However, performing a normalization of the healthy nuclear count against the healthy nuclear count of the DMSO controls (see Materials and Method section) eliminated the uncertainty between mitotic and apoptotic nuclei and corrected the overall information on the healthy nuclear count per well.

These data suggested that the classification of cells based on their nuclear phenotype can be used as a surrogate of more complex gating protocols. Thus, the gating based on the Hoechst33342 signal simplified the High-Content assay setups by not only enabling the counting and identification of cells but also by assessing their health state and compound properties such as intrinsic compound fluorescence or the occurrence of compound precipitation.
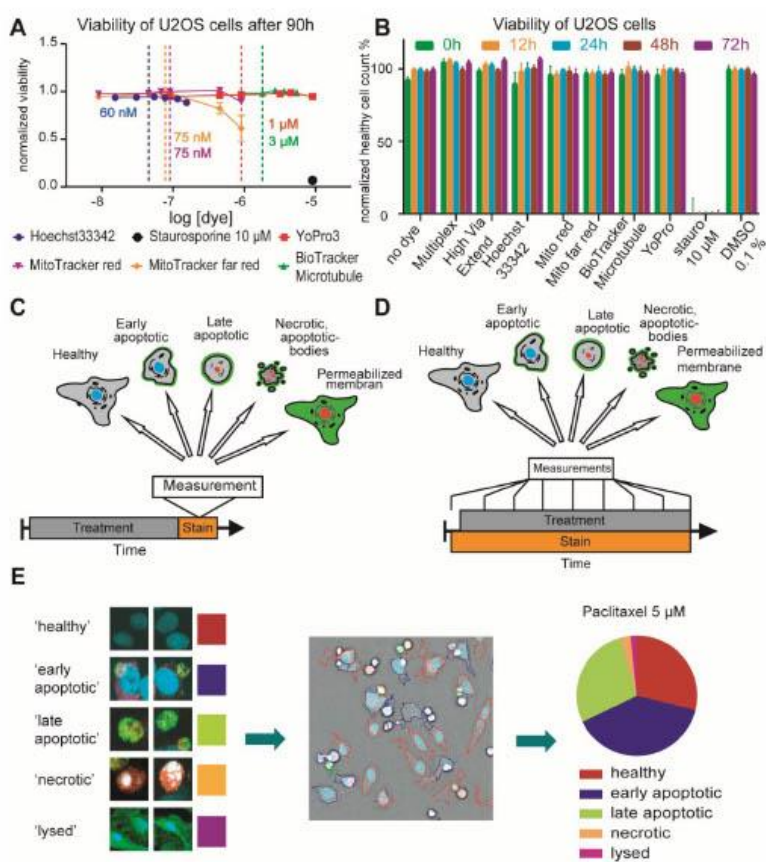

Figure 1. *Cont.*
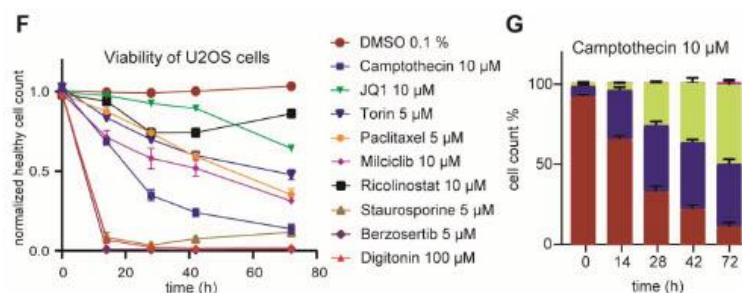
Figure 1. Validation of cell staining dyes and optimization of High-Via Protocol. (A) Healthy cell count after 90 h of cell staining dye exposure at six different concentrations and 10 µM of staurosporine in U2OS cells. Vertical lines show optimal concentration used in subsequent assays (Hoechst33342 60 nM (ThermoFisher, Waltham, MA, USA), MitoTracker ™ red 75 nM (Invitrogen, Waltham, MA, USA), MitoTracker ™ far red 75 nM (Invitrogen, Waltham, MA, USA), YoPro3 1 µM (Invitrogen, Waltham, MA, USA), BioTracker Microtubule 3 µM (EMD Millipore, Burlington, MA, USA). Error bars show SEM of four technical replicates. (B) Healthy cell count 0 h, 12 h, 24 h, 48 h and 72 h after cell staining dye exposure alone or in combination (Multiplex: Hoechst33342 60 nM, BioTracker™ Microtubule 3 µM, MitoTracker ™ red 75 nM, Annexin V Alexa Fluor 680 0.3 µL/well; HighVia Extend: Hoechst33342 60 nM, MitoTracker ™ far red 75 nM, YoPro3 1 µM, Annexin V Alexa Fluor 488 0.3 µL/well) and 10 µM of staurosporine (stauro) normalized to healthy cells exposed to DMSO 0.1% in U2OS cells. Error bars show SEM of four technical replicates. (C) Generic workflow of cell gating in HighVia protocol. (D) Generic workflow of cell gating in HighVia Extend protocol. (E) Cellular classification in healthy, early apoptotic, late apoptotic, necrotic and lysed by trained cell biologist. Cellular classifications shown after segmentation using a machine learning algorithm. Fractions of healthy, early apoptotic, late apoptotic, necrotic and lysed cells after 24 h of 5 µM paclitaxel exposure in U2OS cells. (F) Healthy cell count before and 14 h, 28 h, 42 h and 72 h after compound exposure (camptothecin 10 µM, JQ1 10 µM, torin 5 µM, paclitaxel 5 µM, milciclib 10 µM, ricolinostat 10 µM, staurosporine 5 µM, berzosertib 5 µM, digitonin 100 µM) normalized to healthy cells exposed to DMSO 0.1% in U2OS cells. Error bars show SEM of technical triplicates. (G) Fractions of healthy, early apoptotic, late apoptotic, necrotic and lysed cells before and 14 h, 28 h, 42 h and 72 h after 10 µM camptothecin exposure in U2OS cells. Error bars show SEM of technical triplicates.

2.3. FUCCI Cell Cycle Analysis

Validating effects of small molecules on the cell cycle is an important test for new drug candidates which is frequently assessed, e.g., by Fluorescence-activated cell sorting (FACS) using DNA-binding dyes such as propidium iodine (PI) [33] or the cell cycle dependent degradation of fluorescent maker proteins described by Sakaue-Sawano et al. [34]. The FUCCI technology allows us to distinguish between live cells in different cell phases using a dual-color imaging system [35]. Thanks to the opposing effects of the licensing factor Cdt1 (RFP-tagged) and its inhibitor Geminin (GFP-tagged) on DNA replication, their presence, seen by the fluorescent tag, can be used to distinguish between G1 and S/G2/M phases of the cell cycle. Cells in S/G2/M are identified by a GFP-labeled nucleus (hereafter referred to as "green"). Cells in G1 result in RFP-labeled nuclei ("red") and those in the transition state between G1 and S phase, show both GFP and RFP-labeled nuclei ('yellow"). A small fraction of non-labeled nuclei that appear shortly in between M and G1 phase is rare and can be neglected in the analysis [36].
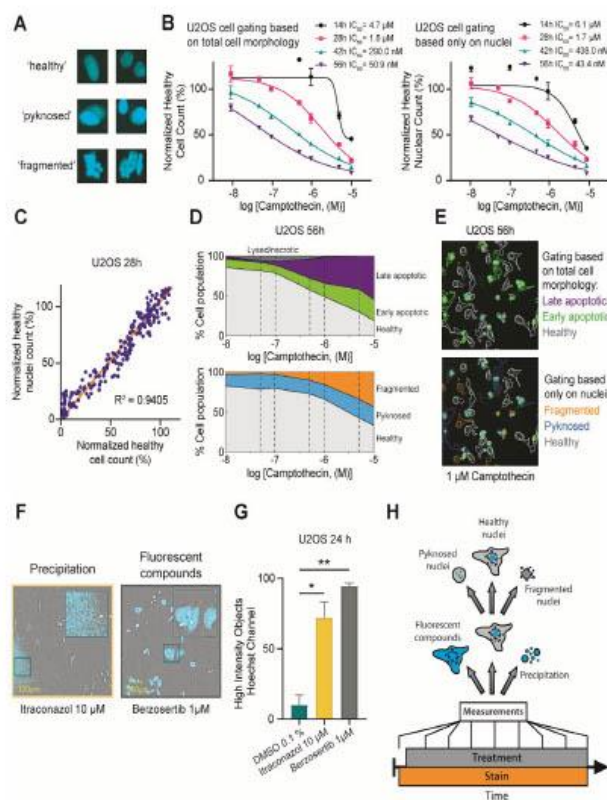
Figure 2. Analysis of Cell Nuclei by Hoechst Channel Intensity level. (**A**) Nuclei classification in healthy, pyknosed and fragmented by trained cell biologist. (**B**) Normalized healthy cell count and normalized healthy nuclear count of different concentrations (0.01 µM, 0.05 µM, 0.1 µM, 0.5 µM, 1 µM, 5 µM, 10 µM) of camptothecin exposure with calculated IC50 values after 14 h, 28 h, 42 h and 56 h. (**C**) Correlation between the healthy cell count and the healthy nuclei count after 28 h of compound exposure normalized to healthy cells exposed to DMSO 0.1% in U2OS cells. (**D**) Fractions of HighVia gating and fractions of healthy, fragmented and pyknosed nuclei after exposure to different concentrations (0.01 µM, 0.05 µM, 0.1 µM, 0.5 µM, 1 µM, 5 µM, 10 µM of camptothecin in U2OS cells after 56 h of compound exposure. Error bars show SEM of three technical replicates. (**E**) Fluorescence image of different gatings of U2OS cells exposed to 1 µM camptothecin after 56 h based on total cell morphology and based on nuclei. (**F**) Bright field confocal image of stained (blue: DNA/nuclei, green: Annexin V apoptosis marker, red: YoPro3, magenta: Mitotracker Deep Red, mitochondria content) U2OS cells after 24 h of compound exposure (itraconazole 10 µM, berzosertib 1 µM). Precipitation of 10 µM itraconazole and fluorescence of 1 µM berzosertib exposure shown as Hoechst High Intensity Objects are highlighted. (**G**) Hoechst High Intensity Objects after 24 h of compound exposure (itraconazole 10 µM, berzosertib 1 µM) and DMSO 0.1% as negative control. Error bars show SEM of three technical replicates. (**H**) Generic workflow of cell gating based on cell nuclei.

To test the compatibility of the nuclei-based gating protocol in combination with other fluorescent markers, we used this cell cycle reporter together with the described nuclear gating strategy in HCT116 cells. We chose this cell line for this experiment due to its favorable nucleus:cell-body distribution.. In our analysis, we focused on the cell cycle phases of unaltered, 'healthy' gated nuclei, but also the pyknosed, and to a certain extent, the fragmented populations could be further gated based on the intensities of the FUCCI reporters (general workflow see Figure 3A). We only considered the three major populations of "green", "red" and "yellow" nuclei, while the neglectable fraction of non-labeled cells was excluded (Figure 3B).

We first assessed whether the introduction of the FUCCI system would influence sensitivity of cells with respect to compound viability. HCT116-FUCCI cells treated with the CDK inhibitor milciclib resulted in a comparable cytotoxic profile as observed with the same protocol above for U2OS cells (Figure 3C). The gating based on the nuclear phenotype enabled the exclusive analysis of cells not showing an apoptotic or damaged phenotype over several time points in one experiment. Comparing the effect of HCT116 treated with 1 µM of milciclib with the one of DMSO treated cells resulted in the expected alteration of cell cycle distribution upon compound treatment. Milciclib treated cells not only showed a lower number of "healthy" classified cells but also displayed a higher population of 'red' nuclei after 18 h of compound treatment (Figure 3D), indicative of a G1 phase arrest, consistent with milciclib's ability of inhibiting cyclin-dependent kinases such as CDK2 [37,38]. The timing of the accumulation of cells in the G1 phase after 18 h correlated well with the less pronounced cytotoxicity at the earliest time point of 3.5 h, further pointing to a primary rather than secondary compound effect. These data underlined the advantages of simultaneously detecting cytotoxicity and investigating different phenotypes at several time points in one experiment.

Encouraged by these results, we analyzed the effect of additional 17 compounds at a single concentration of 10 µM in HCT116-FUCCI cells for up to 70 h (Figure 3E). This set of compounds included compounds known to affect the cell cycle, such as the topoisomerase II inhibitors daunorubicin, doxorubicin as well as mitoxantrone and topoisomerase I inhibitors camptothecin and topotecan. We also included paclitaxel and vinorelbine tartrate, representing drugs that influence the tubulin function. Overall, the compounds represented a broad range of mechanisms affecting the cell cycle, such as triggering checkpoint response, and various cytotoxic agents. A full list of known effects of the used compounds is included in Supplementary Table S3.
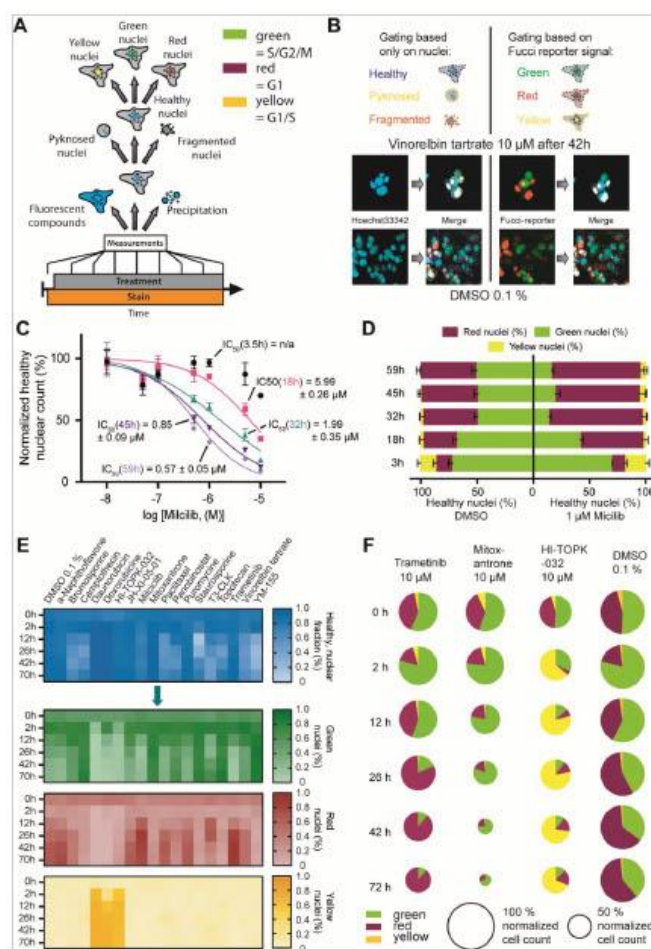
Figure 3. FUCCI Cell Cycle Assay. (**A**) Gating scheme of the machine learning-based analysis of the HCT116-FUCCI reporter cell line (**B**) Example images of HCT116 cells treated either with 10 μM vinorelbine tartrate or 0.1% DMSO for 42 h. Highlighted by colored lines are the gating results of both the nuclear gating step into "healthy", "pyknosed" and "fragmented" nuclei and the FUCCI gating step of 'healthy' nuclei only. (**C**) Cytotoxicity of milciclib for HCT116-FUCCI at different time points assessed by the normalized healthy cell count. The depicted IC50 values are the average and SEM of two independent experiments. (**D**) Population distribution of 'healthy' gated (based on nuclear features) HCT116-FUCCI cells treated with 1 μM milciclib or 0.1% DMSO for different periods. (**E**) Heat map analysis of HCT116-FUCCI cells treated with 10 μM of different inhibitor for different time points. (**F**) Detailed representation of the population fractions of HCT116-FUCCI cells exposed to 10 μM of trametinib, mitoxantrone and HI-TOPK-032. The size of the pie charts reflected the normalized cell count. For comparison, the fractions of DMSO are shown as well.

The population analysis confirmed an increase in red nuclei for the flavone derivate α-naphthoflavone [39], the tubulin binding taxol-derivative paclitaxel [40,41] and MEK1/2 inhibitor trametinib [42] consistent with previous reports of these compounds to cause a cell cycle arrest in G1. An increase in green nuclei, in comparison to DMSO, was detected for both tested topoisomerases I inhibitors, camptothecin and topotecan, as well as mitoxantrone, which are known to cause mitotic cell cycle arrest [27,43–46]. More yellow nuclei were observed for cells treated with the topoisomerase II inhibitors daunorubicin and doxorubicin, both of which are known to cause cell cycle arrest due to DNA double strand breaks [44,47]. Treatment of cells with the TOPK inhibitor HI-TOPK-032 [48] also resulted in an increased number of cells with yellow nuclei, indicating induction of an S phase arrest. In general, all compounds that have been described to interfere in cell cycle progression, showed the expected effects in the FUCCI assay system. Figure 3F shows an example of the nuclei population analysis for the three compounds, trametinib, mitoxantrone and HI-TOPK-032 in comparison to cells exposed to DMSO 0.1%. Additional data can be found in Supplementary Table S4.

2.4. Multiplex Protocol

Low cell viability can be the result of on-target effects, off-target effects or be based on undesirable characteristics of compounds that interfere with the assay system. We therefore included the assessment of tubulin structure, mitochondrial mass changes and membrane permeability in the phenotypical analysis. We used a test set of 21 compounds with known effects on cell viability to validate the protocol (Supplementary Table S6). However, we envisioned that depending on the phenotypes of interest, additional cell staining dyes can be used to detect further changes in cellular morphology or health. Compounds that modulate microtubule functions have been used extensively in cancer research, because of their interference with tumor growth. However, unspecific or unrecognized tubulin binding can lead to false interpretation of presumed target specific effects, in particular in oncological assays. We therefore included a microtubule cytoskeleton dye to detect changes in the tubulin structure. In addition, mitochondrial health was monitored using Mitotracker™ Red (Invitrogen), providing information on mitochondrial mass, which has been shown to provide a good indicator for the apoptotic susceptibility of cells. Higher mitochondrial mass can lead to cell damage, because mitochondria are the main source of ROS (reactive oxygen species) [26]. Membrane permeability was detected using the microtubule cytoskeleton dye and cells were compared to digitonin, as a cell lysis control. Cellular shape and fluorescence were measured as before, at 12 h and 24 h after compound treatment. To analyze effects on tubulin, mitochondria and membrane permeability, a machine learning-based protocol was implemented, based on four earlier tested compounds as a training set for the algorithm (Supplementary Table S6). Cells that showed different tubulin appearance compared to the DMSO 0.1% control were marked as 'tubulin effect'. Cells that showed an increase in mitochondrial mass in comparison to DMSO 0.1% were marked 'mitochondrial mass increase' and cells that showed membrane permeability were marked 'membrane permeable'. For membrane permeability, the reference compound digitonin, a cell detergent leading to cell perforation, was used. In our first approach, we gated all cells, independently of their viability, into the different phenotypical groups, namely 'tubulin effect' or 'no tubulin effect'. With this first analysis, (Supplementary Figure S4) we saw that an increase in mitochondrial mass as well as a tubulin effect was frequently associated with cytotoxicity. When adding 'cytotoxic' compounds such as staurosporine (10 μM) or puromycin (10 μM), most cells showed as expected increased mitochondrial mass and tubulin effects [25]. To exclude these phenotypic effects caused merely by cell death, only cells defined as 'healthy nuclei' were gated into the different groups. The generic workflow of the analysis is shown in Figure 4A. Compounds that have certain quality deficiencies, shown by not passing the property thresholds, were marked. First, compounds that showed a Hoechst High Intensity Object ratio of more than 50% (Figure 4B)

were selected, because they either precipitated or showed auto-fluorescent, interfering with the assay readout.
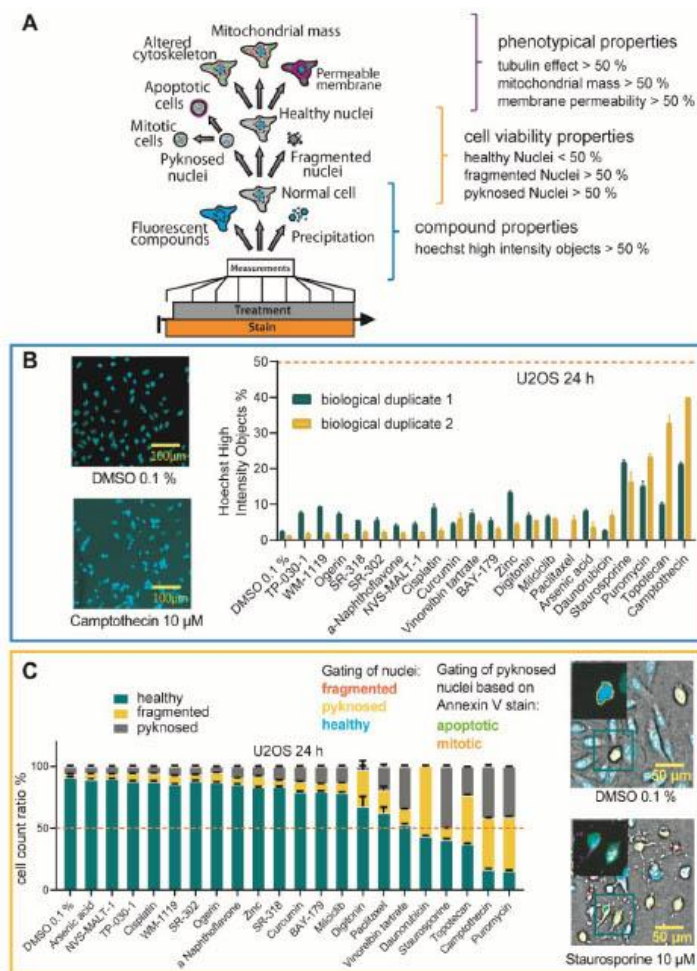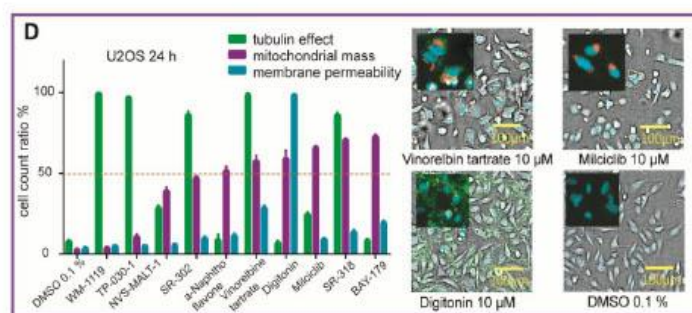

Figure 4. *Cont.*

Figure 4. High Content Multiplex screen of different compounds in U2OS cells. (**A**) General workflow of Multiplex High Via protocol analysis with property thresholds. (**B**) Ratio of Hoechst High Intensity Objects after 24 h of compound exposure (Supplementary Table S5) in U2OS in comparison to DMSO 0.1% of biological duplicates. Error bars show SEM of technical triplicates. Property threshold at 50% marked red. For example, fluorescence confocal images of stained (blue: DNA/nuclei, Hoechst33342) U2OS cells after 24 h of exposure to camptothecin 10 μM in comparison to DMSO 0.1%. (**C**) Cell count ratio of different Nuclei gating after 24 h of 10 μM of compound exposure (Supplementary Table S5) in comparison to DMSO 0.1% in U2OS cells. Error bars show SEM of technical triplicates. Property threshold at 50% marked red. Brightfield confocal image of stained (blue: DNA/nuclei, green: Microtubule different, red: mitochondria content, magenta: Annexin V apoptosis marker) U2OS cells after 24 h of compound exposure of staurosporine 10 μM) in comparison to DMSO 0.1%. Gating of cells for Nuclei gating and Annexin V based gating shown. (**D**) Cell count ratio of tubulin effect (green), mitochondrial mass increase (magenta) and membrane mass permeability (blue) of U2OS cells after 24 h of 10 μM of compound exposure (WM-1119, TP-030-1, NVS-MALT-1, SR-302, α-naphthoflavone, vinorelbine tartrate, digitonin, milciclib, SR-318, BAY-179) in comparison to DMSO 0.1%. Error bars show SEM of technical triplicates. Property threshold at 50% marked. Brightfield confocal image of stained (blue: DNA/nuclei, green: microtubule different, red: mitochondria content, magenta: Annexin V apoptosis marker) U2OS cells after 24 h of compound exposure (vinorelbine tartrate 10 μM, milciclib 10 μM, digitonin 10 μM) in comparison to DMSO 0.1%. Fluorescent image of different staining is highlighted.

For the test set of 21 compounds (Supplementary Table S5), two compounds (camptothecin and topotecan) showed Hoechst High Intensity Objects at levels > 30% after 24 h in one of two biological replicates at 10 μM. Image validation demonstrated, that depending on the location of the precipitated compound, the intensity levels of the channels could vary, thus both duplicates should be considered as precipitation is a stochastic event and dependents on compound handling [49]. All other "normal" cells were then gated based on their nuclei properties in "healthy", "pyknosed" or "fragmented" (Figure 4C). As mentioned above, a distinction between cells that showed condensed nuclei while undergoing apoptosis and cells that were mitotic was only possible taking into account the total cell number and comparison to DMSO as a control. To increase the robustness of this parameter, we added Annexin V as an apoptotic marker to the Hoechst33342 stain. After inclusion of this marker, it was possible to distinguish between mitosis and apoptosis for U2OS cells. Even for the human embryonic kidney cells (HEK293T), which are smaller and rounder than the other cell lines used, the distinction between mitosis and apoptosis was confirmed by normalization to DMSO as control. For the tested small chemogenomic set, all compounds that showed less than 50% of healthy cells were marked and should be assessed further. In our test set, five compounds were marked at 10 μM (daunorubicin, staurosporine, topotecan, camptothecin and puromycin), as expected. To detect phenotypical properties, that are independent of cell death, only nuclei that were gated healthy earlier, were considered. Vinorelbine tartrate, a vinca alkaloid with antimicrotubule

properties interfering with mitotic spindle function, was used as a positive control for tubulin effects [50]. As expected, after 24 h, vinorelbine tartrate treated cells, exhibited increased tubulin effects compared to DMSO 0.1% treated cells. For mitochondrial mass increase, we used milciclib as a test compound. Milciclib, is a known apoptosis modulator [51] and showed an increase in mitochondrial mass over 60% in comparison to DMSO. The chemical probe compounds, SR-318 (a chemical probe for MAPK14) [52] and BAY-179 (a chemical probe for complex I) also showed over 80% healthy nuclei but an increase in mitochondrial mass over 60%. The MAPK14 probe SR-318 as well as the orthogonal dual MAPK14/DDR probe SR-302 also showed a tubulin effect of more than 90%. Interestingly, compounds that are known to permeabilize the membrane such as digitonin (10 μM), showed still more than 50% healthy nuclei after 24 h. The same compounds were tested in HEK293T and MRC-9 cells. The data have been included in the Supplemental Figures S5–S8.

The tested compounds were used, to establish a protocol for detection of multiple readouts. The following requirements were determined for compound flagging: Hoechst High Intensity Objects > 50%, Healthy Nuclei < 50%, Pyknosed Nuclei > 50%, Fragmented Nuclei > 50%, Tubulin effect > 50%, mitochondrial mass increase > 50%, membrane permeability > 50%. However, we recommend that every experiment should contain control compounds with known characteristics as an internal standard. This primary screen was able to flag compounds, which should be further investigated, regarding their suitability for inclusion as compounds in a chemogenomic set. For instance, target specific and off-targeted mechanisms affecting cellular health can be distinguished by control compounds and/or alternative inhibitors with diverse chemical structure that makes it unlikely that these targets also inter with similar off-target mechanisms as the investigated compound.

2.5. Multiplex Analysis of Chemogenomic Compounds

To validate whether this assay can be performed in a medium throughput format, we tested a small library of 215 compounds at two different concentrations, 1 μM and 10 μM, in U2OS, HEK293T and MRC-9 cells. Here, we describe the analysis of the U2OS cells while the results for the further two cell lines are provided in the Supplementary Materials (Supplementary Table S8). Most of the compounds tested have cellular on-target activities in the nanomolar range. It was thus not surprising to observe strong viability effects for a large number of compounds at 10 μM (Figure 5A). We therefore mainly evaluated the effects at the lower concentration of 1 μM. In the first step, compounds that showed Hoechst High Intensity Objects were considered as described earlier. Here, only three of the 215 compounds showed more than 50% of Hoechst intensity compared to normal nuclei after 1 μM treatment: the control compounds berzosertib and camptothecin, as well as the FGFR inhibitor PD173074 [53], which precipitated, a property that can be explained by its hydrophobicity (logP of 4.7). It should be noted that PD173074 has a cellular activity at less than 25 nM and should therefore be used at lower concentrations than 1 μM. The validation of cell viability using the nuclei gating described earlier, revealed 20 compounds with less than 50% of healthy nuclei. Among these, only nine compounds (volasertib, BMS-754807, DDR-TRK-1N, TP0903, GNF-5837, infigratinib, adavosertib, ML154, omipalisib) showed 40% or less healthy nuclei, whereas for example the dual PI3K-AKT-mTOR inhibitor omipalisib, known to cause apoptosis in this concentration range [54], as well as the multi-kinase inhibitor TP0903 [55] and the TRK inhibitor GNF-5837 [56], both known to have an impact on cell viability, decreased the healthy nuclei count more than 60%. Compounds that showed a phenotypic characteristic above the threshold in all three cell lines were 'flagged'. Further investigations are warranted, if the mode of action is responsible for the 'flagged' phenotype or unintended compound features, such as off-target effects or inappropriate concentrations used. In our test chemogenomic set, 49 compounds were 'flagged' (Figure 5B). For example, compound KN-62 was marked due to tubulin effects greater 90%. KN-62 is a calcium/calmodulin kinase inhibitor [57], which inhibits the polymerization of tubulin [57], so the phenotypical effect can be explained by its mode of action. For the TIE2 inhibitor BAY-826, there is no link to tubulin function known so far [58]. However, the recently described off-target activity on DDR1/2 may explain the

observed phenotype [59]. In total, eight compounds were marked to have tubulin effects while they crossed no other phenotypic threshold (BAY-826, bromosporine, CINK4, PF-299804, SR318, SU11274, YM-201636, ZM447439) in U2OS cells. We detected mitochondrial mass increase for 10 compounds (azelastine, GSK1070916, JNJ-5207787, ML-290, NVP-AEW541, PD 102807, SGC-GAK-1, TC-G 1003, topotecan, XMD17-109) whereas only the pan-HER kinase inhibitor PF-299804 [60] demonstrated membrane permeabilization effects of more than 65%. Importantly, the protocol allows for continuous monitoring enabling the detection of time-dependent observations. For example, WZ-4002, a mutant selective covalent EGFR inhibitor showed initially (12 h) at 1 μM an effect on tubulin and mitochondrial mass before causing membrane permeabilization at 24 h.
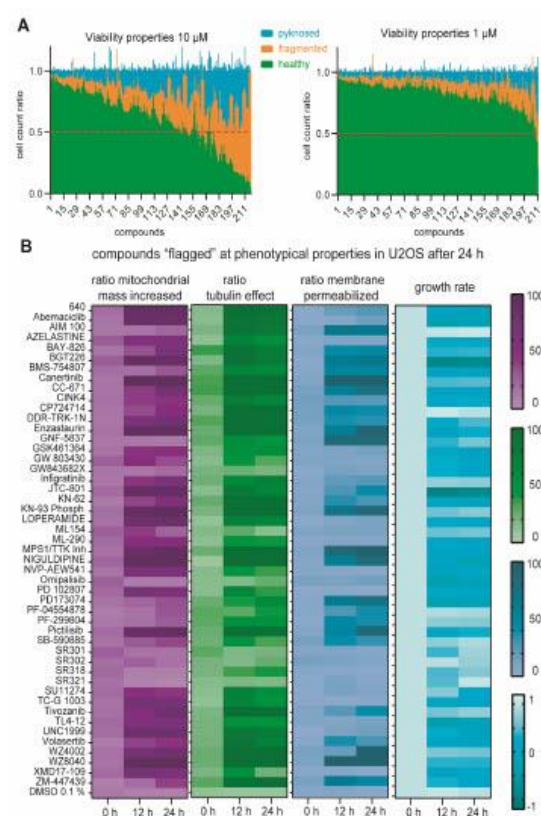
Figure 5. Multiplex Certificates of Chemogenomic Compounds. (**A**) Cell count ratio of different nuclei gating after 24 h of 10 μM and 1 μM of compound exposure (Supplementary Table S8) in U2OS cells. Error bars show SEM of biological duplicates. Healthy nuclei count at 50% (viability threshold) marked as a red line. (**B**) Heat map of phenotypical property ratios (tubulin effect, mitochondrial mass increased and membrane permeabilized) and the growth rate, which was calculated against the non-treated cell number, as described earlier by Hafner et al. [61] of U2OS cells exposed to 49 compounds, that were marked as 'flagged' after Multiplex analysis (phenotypical property threshold > 50%). Heat map shows meaning of two biological duplicates. All data are available in Supplementary Table S8.

3. Discussion

Microscopy-based high-content screening, as a strategy for drug discovery, allows monitoring of multiple phenotypes in a fast and economical way [62]. Phenotypic screening has regained attention in drug discovery in recent years. In comparison to target-based drug discovery methods, phenotypic screening does not rely on the knowledge of a specific target *per se* and works as a tool to address complex relations of poorly understood diseases [5]. Extracting information from biological images collected during phenotypic screening and reducing them to a multidimensional profile, a process called image-based profiling, can be used to identify new disease-associated phenotypes, provide a better understanding about target effects and to predict compound activity, toxicity and mechanism of actions [63]. Here, we present HighVia Extend, a live-cell, expandable, unbiased, image-based profiling assay, suitable for real-time measurements [64]. Similar to HighVia, HighVia Extend is modular in nature, inexpensive and flexible, providing the possibility to add additional fluorescent dyes for further readouts or adaptations for the use in different cell lines. Importantly, the assay is applicable for kinetic measurement for over 72 h and can therefore differentiate between primary target effects and secondary phenotypic results caused by the compound treatment. The lack of kinetic information is a frequent problem in phenotypic screens, which monitor endpoints [65]. Using a single readout, Hoechst33342, to assess cell nuclei, we were able to identify healthy cells with high confidence, which enabled the use of additional stains to detect changes in tubulin appearance and mitochondrial content, respectively. Adding the FUCCI system, additional information regarding compounds affecting the cell cycle could be obtained. However, compared to CellPainting, which uses mostly fixed cells and is based on the generation and evaluation of thousands of features [3] our assay provides comprehensive information about cytotoxicity with considerably less features. Thus, the subsequent data processing is less demanding on bioinformatics capabilities while providing additional kinetic aspects. The modular nature of the assay allows for free combination with other dyes or a pre-screening of compounds with only Hoechst33342 and nuclear gating of the cells to reduce the costs of live-cell dyes. We also successfully combined this experiment with other less complex cytotoxic screens as primary screens, such as proliferation experiments using a plate-reader based readout assessing the metabolic state of cells.

The presented assay offers a suitable annotation for (chemogenomic) libraries, providing information on the effect of these compounds on cellular health. It can be used in combination with assays assessing other aspects of cellular health, such as proteome stress involving protein misfolding and aggregation to better annotate a compound library [66]. Our assay thus helps to distinguish between false-positive or false-negative results of subsequent phenotypic assays [67,68]. False negative results can for example be caused by compounds with low solubility or precipitation of a compound as well as low permeability properties. Poorly soluble compounds can also cause false positive results, which may arise by causing unspecific cell death. Another potential source of false negative data might arise due to the missing expression of certain proteins in the tested cell line. The use of several cell lines in parallel as well as assessing the expression profiles using mRNA sequencing databases can, to a certain extent, offset this bias. Other compounds may cause false positive signals in cell assays due to reactivity of structural groups under applied conditions such as redox effects, complex formation, intrinsic fluorescence, degradation and others [68,69]. In the literature, already a large number of small molecules have been annotated as substances to frequently interfere with different assays [70]. Additional unspecific effects on cellular viability have been described for compounds binding to tubulin, e.g., Gul et al. showed that the preclinical used MTH1 inhibitor TH588 showed decreased tumor growth due to involvement in microtubule spindle regulation instead of the first investigated target effect [19,71]. The assessment of the tubulin modulating properties of compounds in a library can thus provide an alert with respect to the downstream effect on cell viability, which is particularly important for cancer cell biology.

For compounds without specific binding information to a protein as well as for target validation, the assay can provide a simple profile for each compound in a time dependent manner. By comparing the effect on cellular health for compounds targeting the same protein, unspecific effects can be easily detected using further analysis and clustering of results. Testing a well-annotated compound collection can thus be used to identify new biology mechanisms for known targets or even find new target correlations.

3.1. Materials and Methods HighVia Extend Protocol

For testing the High-Via Extend protocol, nine reference compounds (digitonin, torin, ricolinostat, paclitaxel, staurosporine, JQ1, berzosertib, milciclib, camptothecin) with known mode of actions (Supplementary Table S2) were dissolved in DMSO to a concentration of 10 mM. A 7-point serial dilution of every compound were tested in three different cell lines (HEK293T, U2OS, MRC-9). HEK293T (ATCC® CRL-1573™) and U2OS (ATCC®HTB-96™) were cultured in DMEM plus L-Glutamine (High glucose) supplemented by 10% FBS (Gibco, Waltham, MA, USA) and Penicillin/Streptomycin (Gibco). MRC-9 fibroblasts (ATCC® CCL-2™) were cultured in EMEM plus L-Glutamine supplemented by 10% FBS (Gibco) and Penicillin/Streptomycin (Gibco). Cells were seeded at a density of 1250 cells per well in a 384 well plates in culture medium (Cell culture microplate, PS, f-bottom, µClear®, 781091, Greiner), with a volume of 50 µL per well. All outer wells were filled with 100 µL PBS-buffer (Gibco). Simultaneously with seeding, cells were stained with 60 nM Hoechst33342 (Thermo Scientific, Waltham, MA, USA), 75 nM MitoTracker™ far red (Invitrogen, Waltham, MA, USA), 0.3 µL/well Annexin V Alexa Fluor 488 conjugate (Invitrogen, MA, USA) and 1 µM YoPro3 (Invitrogen, MA, USA).

Cellular shape and fluorescence of the untreated cells was measured 24 h after seeding, using the CQ1 high-content confocal microscope (Yokogawa, Tokyo, Japan). The compounds were added in a 1:1000 dilution (50 nL/well) using an Echo 550 (LabCyte, San Josef, CA, USA) and measured again once and then every 12 h over 72 h. The following setup parameters were used for image acquisition: Ex 405 nm/Em 447/60 nm, 500 ms, 50%; Ex 561 nm/Em 617/73 nm, 100 ms, 40%; Ex 488/Em 525/50 nm, 50 ms, 40%; Ex 640 nm/Em 685/40, 50 ms, 20%; bright field, 300 ms, 100% transmission, one centered field per well, seven z-stacks per well with 55 µm spacing.

Images were analyzed using the CellPathfinder software (Yokogawa), segmented and classified as described previously [22]. Briefly, using an automated algorithm, cell "nuclei" were identified by Hoechst channel intensity levels and optimized by smoothing of mean intensity levels, thresholding and afterwards size-filtering to accurately segment nuclei from cytosol. The 'cell body' was defined using the bright field channel. The digital phase contrast was determined between z-stack 3 and 5 with a phase-contrast level of 0.003 to improve cellular shape separation from background. To better identify cytoplasmic areas, the threshold results of the cell body were defined as interdependent of nuclei. After segmentation of the cells, classification was performed using the machine learning feature of the CellPathfinder Software. Training of the machine learning algorithm was performed by an experienced cell biologist. The cells were classified in healthy, early apoptotic, late apoptotic, necrotic and lysed cells by 19 features of the cell body and 13 features of the nuclei (Supplementary Figure S7 including dye intensity levels and cellular morphology characteristics such as cell diameter or compactness. Different control compounds were used to train the machine learning algorithm. Staurosporine 10 µM was used to identify apoptotic cells, digitonin 10 µM was used to classify lysed cells. The analysis was validated using duplicate wells of the named compounds. For nuclei classification, the cells were subdivided in healthy, pyknosed and fragmented nuclei by ten features (Supplementary Figure S7) of the Hoechst channel. To detect objects that show high intensity of the Hoechst channel, classification in High Intensity Objects and Normal Intensity Objects was implemented using three features for the cell body and two features for the nuclei (Supplementary Figure S7). The health cell count and the healthy nuclei count were normalized against the healthy cell count and healthy nuclei count of cells treated with

0.1% DMSO. Significance was calculated using a two-way ANOVA analysis in GraphPad PRISM 8.

3.2. Multiplex Protocol

HEK293T (ATCC® CRL-1573™) and U2OS (ATCC®HTB-96™) were cultured in DMEM plus L-Glutamine (High glucose) supplemented by 10% FBS (Gibco) and Penicillin/Streptomycin (Gibco). MRC-9 fibroblasts (ATCC® CCL-2™) were cultured in EMEM plus L-Glutamine supplemented by 10% FBS (Gibco) and Penicillin/Streptomycin (Gibco). One day prior to compound exposure, cells were stained simultaneously to seeding with 60 nM Hoechst33342 (Thermo Scientific, MA, USA), 75 nM Mitotracker red (Invitrogen, MA, USA), 0.3 µL/well Annexin V Alexa Fluor 680 conjugate (Invitrogen, MA, USA) and 25 nL/well BioTracker™ 488 Green Microtubule Cytoskeleton Dye (EMD Millipore, MA, USA). Cells were seeded at a density of 2000 cells per well in a 384 well plates in culture medium (Cell culture microplate, PS, f-bottom, µClear®, 781091, Greiner, Frickenhausen, Germany), with a volume of 50 µL per well. All outer wells were filled with 100 µL PBS-buffer (Gibco).

Using the CQ1 high-content confocal microscope (Yokogawa), cellular shape and fluorescence was measured before and 12 h as well as 24 h after compound treatment. All compounds were diluted in DMSO to a concentration of 10 mM. Compounds were added directly to the cells in a 1:1000 dilution (50 nL/well) using an Echo 550 (LabCyte, San Josef, CA, USA).

For image acquisition, the following parameters were used: Ex 405 nm/Em 447/60 nm, 500 ms, 50%; Ex 561 nm/Em 617/73 nm, 100 ms, 40%; Ex 488/Em 525/50 nm, 50 ms, 40%; Ex 640 nm/Em 685/40, 50 ms, 20%; bright field, 300 ms, 100% transmission, one centered field per well, seven z-stacks per well with a total of 55 µm spacing. The rather large spacing distance was used to create a robust readout, compensating potential plate variations and enabling automated screening without the use of autofocus. The overlap of the fluorescence emission spectra of the dyes was neglectable for all but the MitoTracker Red and Annexin V Alexa Fluor 680 (Supplementary Figure S9). However, this overlap does not influence the analysis, since the excitation maxima of these two dyes are well separated and the gating algorithm analyses only the Mitotracker Red intensity in Annexin 5 negative cells.

All images were analyzed using the CellPathfinder software (Yokogawa). Segmentation of cells was performed as described earlier. First, the cells are classified in Hoechst High Intensity Objects or Normal Intensity Objects (Supplementary Table S7). All normal gated cells are further classified in healthy, fragmented or pyknosed nuclei (Supplementary Table S7). The pyknosed cells are gated in mitotic or apoptotic cells using seven features for the cell body and five features for the cell nuclei according to their Annexin V staining intensity (Supplementary Table S7). All cells that were classified as including a healthy nucleus are further gated into three phenotypic classes. They are gated in tubulin effect or no tubulin effect (Supplementary Table S7), mitochondrial mass increased or not increased (Supplementary Table S7) and membrane permeability/membrane normal (Supplementary Table S7). Growth rate was calculated against non-treated cells and cells treated with DMSO 0.1% [61].

3.3. FUCCI Assay Protocol

For generation of a stable cell line, including the florescent ubiquitination-base cell cycle indicator FUCCI, the plasmid-based transposon system Sleeping Beauty was used. Vector (pSBbi_Fucci) and the Transposase SB100X have been described previously [72]. HCT116 cells (ATCC® CCL-247™) were cotransfected in a small cell culture flask (5 mL) with a mixture of 9.5 µg pSBbi_Fucci vector and 0.5 µg if the SB100X transposase vector using FuGENE HD (Promega) as described previously [73]. Two days after transfection, cells were selected over 10 days using puromycin (1 µg/mL) and cultivated afterwards for two more weeks in McCoys 5A plus L-Glutamine (Gibco) supplemented by 10% FBS (Gibco) and Penicillin/Streptomycin (Gibco).

HCT116-FUCCI cells were seeded at a density of 1250 cells per well in a 384 well plate (Cell culture microplate, PS, f-bottom, µClear®, 781091, Greiner) in culture medium

to 50 µL per well and stained additionally with 60 nM Hoechst33342 (Thermo Scientific). Outer wells were filled with 100 µL PBS-buffer (Gibco). Fluorescence and cellular shape were measured before and after compound treatment for 72 h every 12 h using the CQ1 high-content confocal microscope (Yokogawa). Compounds were added directly to the cells, using an Echo 550 (LabCyte, San Josef, CA, USA) in a 1:1000 dilution (50 nL/well) to a final concentration of 10 µM.

Following parameters were used for image acquisition: Ex 405 nm/Em 447/60 nm, 500 ms, 50%; Ex 561 nm/Em 617/73 nm, 100 ms, 40%; Ex 488/Em 525/50 nm, 50 ms, 40%; Ex 640 nm/Em 685/40, 50 ms, 20%; bright field, 300 ms, 100% transmission, one centered field per well, seven z-stacks per well with a total of 55 µm spacing. Image analysis was performed using the CellPathfinder software (Yokogawa). Segmentation of cells was performed as described earlier. First, the cells are classified in Hoechst High Intensity Objects or Normal Intensity Objects (Supplementary Table S7). All normal gated cells are further classified in healthy, fragmented or pyknosed nuclei (Supplementary Table S7). The cells that showed healthy nuclei were then further gated in red/green or yellow using 11 features of the cell body and four features of the cell nuclei (Supplementary Figure S7). Total cell count was normalized against total cell count of cells treated with 0.1% DMSO.

3.4. Dye Titration CQ1 and Alamarblue Assay

U2OS cells (ATCC®HTB-96™) were cultured in DMEM plus L-Glutamine (High glucose) supplemented by 10% FBS (Gibco) and Penicillin/Streptomycin (Gibco) and seeded on a 384 well plate in culture medium (Cell culture microplate, PS, f-bottom, µClear®, 781091, Greiner, Frickenhausen, Germany) with half of the plate with and the other half of the plate, without dyes at a density of 1500 cells per well in a, in a volume of 50 µL per well. All outer wells were filled with 100 µL PBS-buffer (Gibco). Then, 24 h after seeding, cells without dyes were treated with the same concentration of dyes and directly measured by the CQ1 confocal microscope (Yokogawa) over 72 h every 12 h. The dyes Hoechst33342 (Thermo Scientific, MA, USA), Mitotracker red (Invitrogen, MA, USA), Mitotracker far red (Invitrogen, MA, USA), Annexin V Alexa Fluor 680 conjugate (Invitrogen, MA, USA), Annexin V Alexa Fluor 488 conjugate (Invitrogen, MA, USA), BioTracker™ 488 Green Microtubule Cytoskeleton Dye (EMD Millipore, Massachusetts, United States), YoPro3 (Invitrogen, MA, USA)) were added using an Echo 550 (LabCyte, San Josef, CA, USA) in a 7-fold dilution and different dye combinations (Supplementary Table S1). Image acquisition was completed with the following parameters: Ex 405 nm/Em 447/60 nm, 500 ms, 50%; Ex 561 nm/Em 617/73 nm, 100 ms, 40%; Ex 488/Em 525/50 nm, 50 ms, 40%; Ex 640 nm/Em 685/40, 50 ms, 20%; bright field, 300 ms, 100% transmission, one centered field per well, seven z-stacks per well with a total of 55 µm spacing. Image analysis was performed using the CellPathfinder software (Yokogawa) as described earlier. To detect the cells without Hoechst33342 stain, the cell body was defined just by bright field intensity levels. Cells were classified using machine learning algorithms by an experienced cell biologist as healthy or not healthy.

After 72 h, the plate was treated with 1:10 alamarBlue™ (ThermoFisher, MA, USA) solution for 12 h. Fluorescence was measured on a PHERAstar plate reader (BMG Labtech, Ortenberg, Germany) with an emission of 590 nm and excitation of 545 nm.

Table S4: Additional data FUCCI assay (Excel). Table S8: Multiplex Chemogenomic data (Excel). References [74–82] are cited in the Supplementary Materials.

Author Contributions: Conceptualization, A.T., M.S. and S.M.; methodology and formal analysis, A.T., M.S. and E.K.; validation A.T., M.S., E.K. and A.C.; data curation, A.T. and M.S.; writing—original draft preparation, A.T., M.S. and S.M.; writing, review and editing, A.T., A.C., E.K., R.M., S.K., M.S. and S.M.; supervision, M.S., R.M., S.K. and S.M.; project administration, S.K. and S.M.; funding acquisition, S.K. and S.M. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: All data used in this study is available in the BioImage Archive under https://www.ebi.ac.uk/biostudies/studies/S-BIAD145 (accessed on 2 January 2022).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Sample Availability: Samples of the compounds are available from the authors.

References

1. Haasen, D.; Schopfer, U.; Antczak, C.; Guy, C.; Fuchs, F.; Selzer, P. How Phenotypic Screening Influenced Drug Discovery: Lessons from Five Years of Practice. *ASSAY Drug Dev. Technol.* **2017**, *15*, 239–246. [CrossRef]
2. Rietdijk, J.; Tampere, M.; Pettke, A.; Georgiev, P.; Lapins, M.; Warpman-Berglund, U.; Spjuth, O.; Puumalainen, M.-R.; Carreras-Puigvert, J. A phenomics approach for antiviral drug discovery. *BMC Biol.* **2021**, *19*, 156. [CrossRef]
3. Bray, M.-A.; Singh, S.; Han, H.; Davis, C.T.; Borgeson, B.; Hartland, C.; Kost-Alimova, M.; Gustafsdottir, S.M.; Gibson, C.C.; Carpenter, A. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* **2016**, *11*, 1757–1774. [CrossRef] [PubMed]
4. Schiff, L.; Migliori, B.; Chen, Y.; Carter, D. Deep learning and automated Cell Painting reveal Parkinson's disease-specific signatures in primary patient fibroblasts. *bioRxiv* **2020**. [CrossRef]
5. Moffat, J.G.; Vincent, F.; Lee, J.A.; Eder, J.; Prunotto, M. Opportunities and challenges in phenotypic drug discovery: An industry perspective. *Nat. Rev. Drug Discov.* **2017**, *16*, 531–543. [CrossRef] [PubMed]
6. Arrowsmith, C.; Audia, J.; Austin, C.; Baell, J.; Bennett, J.; Blagg, J.; Bountra, C.; Brennan, P.; Brown, P.; Bunnage, M.E.; et al. The promise and peril of chemical probes. *Nat. Chem. Biol.* **2015**, *11*, 536–541. [CrossRef] [PubMed]
7. Brown, P.J.; Müller, S. Open access chemical probes for epigenetic targets. *Futur. Med. Chem.* **2015**, *7*, 1901–1917. [CrossRef]
8. Drewes, G.; Knapp, S. Chemoproteomics and Chemical Probes for Target Discovery. *Trends Biotechnol.* **2018**, *36*, 1275–1286. [CrossRef]
9. Bunnage, M.E.; Chekler, E.L.P.; Jones, L. Target validation using chemical probes. *Nat. Chem. Biol.* **2013**, *9*, 195–199. [CrossRef]
10. Wells, C.I.; Al-Ali, H.; Andrews, D.M.; Asquith, C.R.M.; Axtman, A.D.; Dikic, I.; Ebner, D.; Ettmayer, P.; Fischer, C.; Frederiksen, M.; et al. The Kinase Chemogenomic Set (KCGS): An Open Science Resource for Kinase Vulnerability Identification. *Int. J. Mol. Sci.* **2021**, *22*, 566. [CrossRef]
11. Canham, S.M.; Wang, Y.; Cornett, A.; Auld, D.S.; Baeschlin, D.K.; Patoor, M.; Skaanderup, P.R.; Honda, A.; Llamas, L.; Wendel, G.; et al. Systematic Chemogenetic Library Assembly. *Cell Chem. Biol.* **2020**, *27*, 1124–1129. [CrossRef]
12. Dafniet, B.; Cerisier, N.; Boezio, B.; Clary, A.; Ducrot, P.; Dorval, T.; Gohier, A.; Brown, D.; Audouze, K.; Taboureau, O. Development of a chemogenomics library for phenotypic screening. *J. Chemin.* **2021**, *13*, 91. [CrossRef]
13. Müller, S.; Ackloo, S.; Arrowsmith, C.H.; Bauser, M.; Baryza, J.L.; Blagg, J.; Boettcher, J.; Bountra, C.; Brown, P.; Bunnage, M.; et al. Donated chemical probes for open science. *eLife* **2018**, *7*, 7. [CrossRef]

14. Bredel, M.; Jacoby, E. Chemogenomics: An emerging strategy for rapid target and drug discovery. *Nat. Rev. Genet.* 2004, 5, 262–275. [CrossRef]

15. Jones, L.; Bunnage, M.E. Applications of chemogenomic library screening in drug discovery. *Nat. Rev. Drug Discov.* 2017, 16, 285–296. [CrossRef]

16. Caron, P.R.; Mullican, M.D.; Mashal, R.D.; Wilson, K.P.; Su, M.S.; Murcko, M. Chemogenomic approaches to drug discovery. *Curr. Opin. Chem. Biol.* 2001, 5, 464–470. [CrossRef]

17. >EUbOPEN. Available online: https://www.eubopen.org/ (accessed on 5 January 2022).

18. Carter, A.J.; Kraemer, O.; Zwick, M.; Mueller-Fahrnow, A.; Arrowsmith, C.H.; Edwards, A.M. Target 2035: Probing the human proteome. *Drug Discov. Today* 2019, 24, 2111–2115. [CrossRef]

19. Kawamura, T.; Kawatani, M.; Muroi, M.; Kondoh, Y.; Futamura, Y.; Aono, H.; Tanaka, M.; Honda, K.; Osada, H. Proteomic profiling of small-molecule inhibitors reveals dispensability of MTH1 for cancer cell survival. *Sci. Rep.* 2016, 6, 26521. [CrossRef]

20. Sun, H.; Wang, Y.; Cheff, D.M.; Hall, M.D.; Shen, M. Predictive models for estimating cytotoxicity on the basis of chemical structures. *Bioorg. Med. Chem.* 2020, 28, 115422. [CrossRef]

21. Tang, H.; Duggan, S.; Richardson, P.L.; Marin, V.; Warder, S.E.; McLoughlin, S.M. Target Identification of Compounds from a Cell Viability Phenotypic Screen Using a Bead/Lysate-Based Affinity Capture Platform. *J. Biomol. Screen.* 2015, 21, 201–211. [CrossRef]

22. Howarth, A.; Schröder, M.; Montenegro, R.C.; Drewry, D.H.; Sailem, H.; Millar, V.; Müller, S.; Ebner, D.V. HighVia—A Flexible Live-Cell High-Content Screening Pipeline to Assess Cellular Toxicity. *SLAS Discov. Adv. Sci. Drug Discov.* 2020, 25, 801–811. [CrossRef]

23. Chen, A.Y.; Yu, C.; Bodley, A.; Peng, L.F.; Liu, L. A new mammalian DNA topoisomerase I poison Hoechst 33342: Cytotoxicity and drug resistance in human cell cultures. *Cancer Res.* 1993, 53, 1332–1337. [PubMed]

24. Durand, R.E.; Olive, P.L. Cytotoxicity, Mutagenicity and DNA damage by Hoechst 33342. *J. Histochem. Cytochem.* 1982, 30, 111–116. [CrossRef]

25. Camilleri-Broët, S.; Vanderwerff, H.; Caldwell, E.; Hockenbery, D. Distinct Alterations in Mitochondrial Mass and Function Characterize Different Models of Apoptosis. *Exp. Cell Res.* 1998, 239, 277–292. [CrossRef] [PubMed]

26. Márquez-Jurado, S.; Díaz-Colunga, J.; das Neves, R.P.; Martinez-Lorente, A.; Almazan, F.; Guantes, R.; Iborra, F.J. Mitochondrial levels determine variability in cell death by modulating apoptotic gene expression. *Nat. Commun.* 2018, 9, 389. [CrossRef] [PubMed]

27. Hsiang, Y.H.; Hertzberg, R.; Hecht, S.; Liu, L. Camptothecin induces protein-linked DNA breaks via mammalian DNA topoisomerase I. *J. Biol. Chem.* 1985, 260, 14873–14878. [CrossRef]

28. Da Motta, L.L.; Ledaki, I.; Purshouse, K.; Haider, S.; De Bastiani, M.A.; Baban, D.; Morotti, M.; Steers, G.; Wigfield, S.; Bridges, E.; et al. The BET inhibitor JQ1 selectively impairs tumour response to hypoxia and downregulates CA9 and angiogenesis in triple negative breast cancer. *Oncogene* 2017, 36, 122–132. [CrossRef] [PubMed]

29. Francipane, M.G.; Lagasse, E. Selective targeting of human colon cancer stem-like cells by the mTOR inhibitor Torin-1. *Oncotarget* 2013, 4, 1948–1962. [CrossRef] [PubMed]

30. Styrt, B.; Johnson, P.C.; Klempner, M.S. Differential lysis of plasma membranes and granules of human neutrophilis by digitonin. *Tissue Cell* 1985, 17, 793–800. [CrossRef]

31. Wen, N.; Guo, B.; Zheng, H.; Xu, L.; Liang, H.; Wang, Q.; Wang, D.; Chen, X.; Zhang, S.; Li, Y.; et al. Bromodomain inhibitor jq1 induces cell cycle arrest and apoptosis of glioma stem cells through the VEGF/PI3K/AKT signaling pathway. *Int. J. Oncol.* 2019, 55, 879–895. [CrossRef] [PubMed]

32. Vogl, D.T.; Raje, N.; Jagannath, S.; Richardson, P.; Hari, P.; Orlowski, R.; Supko, J.G.; Tamang, D.; Yang, M.; Jones, S.S.; et al. Ricolinostat, the First Selective Histone Deacetylase 6 Inhibitor, in Combination with Bortezomib and Dexamethasone for Relapsed or Refractory Multiple Myeloma. *Clin. Cancer Res.* 2017, 23, 3307–3315. [CrossRef] [PubMed]

33. Kang, H.; Alvarado, A.S. Flow cytometry methods for the study of cell-cycle parameters of planarian stem cells. *Dev. Dyn.* 2009, 238, 1111–1117. [CrossRef] [PubMed]

34. Sakaue-Sawano, A.; Kurokawa, H.; Morimura, T.; Hanyu, A.; Hama, H.; Osawa, H.; Kashiwagi, S.; Fukami, K.; Miyata, T.; Miyoshi, H.; et al. Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. *Cell* 2008, 132, 487–498. [CrossRef]

35. Yano, S.; Hoffman, R.M. Real-Time Determination of the Cell-Cycle Position of Individual Cells within Live Tumors Using FUCCI Cell-Cycle Imaging. *Cells* 2018, 7, 168. [CrossRef]

36. Zielke, N.; Edgar, B.A. FUCCI sensors: Powerful new tools for analysis of cell proliferation. *Wiley Interdiscip. Rev. Dev. Biol.* 2015, 4, 469–487. [CrossRef]

37. Jorda, R.; Hendrychová, D.; Voller, J.; Řezníčková, E.; Gucký, T.; Kryštof, V. How Selective Are Pharmacological Inhibitors of Cell-Cycle-Regulating Cyclin-Dependent Kinases? *J. Med. Chem.* 2018, 61, 9105–9120. [CrossRef] [PubMed]

38. Sánchez-Martínez, C.; Gelbert, L.M.; Lallena, M.J.; de Dios, A. Cyclin dependent kinase (CDK) inhibitors as anticancer drugs. *Bioorg. Med. Chem. Lett.* 2015, 25, 3420–3435. [CrossRef]

39. Reiners, J.J.; Clift, R.; Mathieu, P. Suppression of cell cycle progression by flavonoids: Dependence on the aryl hydrocarbon receptor. *Carcinogenesis* 1999, 20, 1561–1566. [CrossRef] [PubMed]

40. Demidenko, Z.N.; Kalurupalle, S.; Hanko, C.; Lim, C.-U.; Broude, E.; Blagosklonny, M.V. Mechanism of G1-like arrest by low concentrations of paclitaxel: Next cell cycle p53-dependent arrest with sub G1 DNA content mediated by prolonged mitosis. *Oncogene* 2008, 27, 4402–4410. [CrossRef] [PubMed]

41. Wang, T.-H.; Wang, H.-S.; Soong, Y.-K. Paclitaxel-induced cell death. *Cancer* 2000, 88, 2619–2628. [CrossRef]
42. Schick, U.; Kyula, J.; Barker, H.; Patel, R.; Zaidi, S.; Gregory, C.; Hafsi, H.; Roulstone, V.; Deutsch, E.; McLaughlin, M.; et al. Trametinib radiosensitises RAS- and BRAF-mutated melanoma by perturbing cell cycle and inducing senescence. *Radiother. Oncol.* 2015, 117, 364–375. [CrossRef] [PubMed]
43. Faulds, D.; Balfour, J.A.; Chrisp, P.; Langtry, H.D. Mitoxantrone. *Drugs* 1991, 41, 400–449. [CrossRef] [PubMed]
44. Kluza, J.; Marchetti, P.; Gallego, M.-A.; Lancel, S. Mitochondrial proliferation during apoptosis induced by anticancer agents: Effects of doxorubicin and mitoxantrone on cancer and cardiac cells. *Oncogene* 2004, 23, 7018–7030. [CrossRef] [PubMed]
45. Pommier, Y. Topoisomerase I inhibitors: Camptothecins and beyond. *Nat. Cancer* 2006, 6, 789–802. [CrossRef]
46. Tsunetoh, S.; Terai, Y.; Sasaki, H.; Tanabe, A.; Tanaka, Y.; Sekijima, T.; Fujioka, S.; Kawaguchi, H.; Kanemura, M.; Yamashita, Y.; et al. Topotecan as a molecular targeting agent which blocks the Akt and VEGF cascade in platinum-resistant ovarian cancers. *Cancer Biol. Ther.* 2010, 10, 1137–1146. [CrossRef] [PubMed]
47. Al-Aamri, H.M.; Ku, H.; Irving, H.R.; Tucci, J.; Meehan-Andrews, T.; Bradley, C. Time dependent response of daunorubicin on cytotoxicity, cell cycle and DNA repair in acute lymphoblastic leukaemia. *BMC Cancer* 2019, 19, 1–12. [CrossRef]
48. Ishikawa, C.; Senba, M.; Mori, N. Mitotic kinase PBK/TOPK as a therapeutic target for adult T-cell leukemia/lymphoma. *Int. J. Oncol.* 2018, 53, 801–814. [CrossRef]
49. Delaney, J.S. Predicting aqueous solubility from structure. *Drug Discov. Today* 2005, 10, 289–295. [CrossRef]
50. Knick, V.C.; Eberwein, D.J.; Miller, C.G. Vinorelbine Tartrate and Paclitaxel Combinations: Enhanced Activity Against In Vivo P388 Murine Leukemia Cells. *J. Natl. Cancer Inst.* 1995, 87, 1072–1077. [CrossRef] [PubMed]
51. Johansson, P.; Krona, C.; Kundu, S.; Doroszko, M.; Baskaran, S.; Schmidt, L.; Vinel, C.; Almstedt, E.; Elgendy, R.; Elfineh, L.; et al. A Patient-Derived Cell Atlas Informs Precision Targeting of Glioblastoma. *Cell Rep.* 2020, 32, 107897. [CrossRef] [PubMed]
52. Röhm, S.; Berger, B.-T.; Schröder, M.; Chaikuad, A.; Winkel, R.; Hekking, K.F.W.; Benninghof, J.J.C.; Mueller, G.; Tesch, R.; Kudolo, M.; et al. Fast Iterative Synthetic Approach toward Identification of Novel Highly Selective p38 MAP Kinase Inhibitors. *J. Med. Chem.* 2019, 62, 10757–10782. [CrossRef] [PubMed]
53. Pardo, O.; Latigo, J.; Jeffery, R.E.; Nye, E.; Poulsom, R.; Spencer-Dene, B.; Lemoine, N.; Stamp, G.W.; Aboagye, E.; Seckl, M.J. The Fibroblast Growth Factor Receptor Inhibitor PD173074 Blocks Small Cell Lung Cancer Growth In Vitro and In Vivo. *Cancer Res.* 2009, 69, 8645–8651. [CrossRef]
54. Ippolito, T.; Tang, G.; Mavis, C.; Gu, J.J.; Hernandez-Ilizaliturri, F.J.; Barth, M.J. Omipalisib (GSK458), a Novel Pan-PI3K/mTOR Inhibitor, Exhibits In Vitro Anti-Lymphoma Activity in Chemotherapy-Sensitive and -Resistant Models of Burkitt Lymphoma. *Blood* 2016, 128, 5376. [CrossRef]
55. Aveic, S.; Corallo, D.; Porcù, E.; Pantile, M.; Boso, D.; Zanon, C.; Viola, G.; Sidarovich, V.; Mariotto, E.; Quattrone, A.; et al. TP-0903 inhibits neuroblastoma cell growth and enhances the sensitivity to conventional chemotherapy. *Eur. J. Pharmacol.* 2018, 818, 435–448. [CrossRef]
56. Chen, Y.; Wang, H.; Chen, Y.; Wang, M.; Ding, G.; Li, T. Trk inhibitor GNF-5837 suppresses the tumor growth, survival and migration of renal cell carcinoma. *Oncol. Rep.* 2019, 42, 2039–2048. [CrossRef] [PubMed]
57. Williams, E.J.; Mittal, B.; Walsh, F.S.; Doherty, P. A Ca2+/Calmodulin Kinase Inhibitor, KN-62, Inhibits Neurite Outgrowth Stimulated by CAMs and FGF. *Mol. Cell. Neurosci.* 1995, 6, 69–79. [CrossRef] [PubMed]
58. Schneider, H.; Szabo, E.; Machado, R.A.C.; Broggini-Tenzer, A.; Walter, A.; Lobell, M.; Heldmann, D.; Süssmeier, F.; Grünewald, S.; Weller, M. Novel TIE-2 inhibitor BAY-826 displays in vivo efficacy in experimental syngeneic murine glioma models. *J. Neurochem.* 2016, 140, 170–182. [CrossRef]
59. Lino, M.; Wan, M.H.; Rocca, A.S.; Ngai, D.; Shobeiri, N.; Hou, G.; Ge, C.; Franceschi, R.T.; Bendeck, M.P. Diabetic Vascular Calcification Mediated by the Collagen Receptor Discoidin Domain Receptor 1 via the Phosphoinositide 3-Kinase/Akt/Runt-Related Transcription Factor 2 Signaling Axis. *Arter. Thromb. Vasc. Biol.* 2018, 38, 1878–1889. [CrossRef] [PubMed]
60. Reckamp, K.L.; Giaccone, G.; Camidge, D.R.; Gadgeel, S.M.; Khuri, F.R.; Engelman, J.A.; Koczywas, M.; Rajan, A.; Campbell, A.K.; Gernhardt, D.; et al. A phase 2 trial of dacomitinib (PF-00299804), an oral, irreversible pan-HER (human epidermal growth factor receptor) inhibitor, in patients with advanced non–small cell lung cancer after failure of prior chemotherapy and erlotinib. *Cancer* 2014, 120, 1145–1154. [CrossRef]
61. Hafner, M.; Niepel, M.; Chung, M.; Sorger, P.K. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nat. Methods* 2016, 13, 521–527. [CrossRef] [PubMed]
62. Boutros, M.; Heigwer, F.; Laufer, C. Microscopy-Based High-Content Screening. *Cell* 2015, 163, 1314–1325. [CrossRef] [PubMed]
63. Chandrasekaran, S.N.; Ceulemans, H.; Boyd, J.D.; Carpenter, A.E. Image-based profiling for drug discovery: Due for a machine-learning upgrade? *Nat. Rev. Drug Discov.* 2021, 20, 145–159. [CrossRef]
64. Cole, R. Live-cell imaging. *Cell Adhes. Migr.* 2014, 8, 452–459. [CrossRef]
65. Neumann, B.; Held, M.; Liebel, U.; Erfle, H.; Rogers, P.; Pepperkok, R.; Ellenberg, J. High-throughput RNAi screening by time-lapse imaging of live human cells. *Nat. Methods* 2006, 3, 385–390. [CrossRef]
66. Liu, Y.; Fares, M.; Dunham, N.P.; Gao, Z.; Miao, K.; Jiang, X.; Bollinger, S.S.; Boal, A.K.; Zhang, X. AgHalo: A Facile Fluorogenic Sensor to Detect Drug-Induced Proteome Stress. *Angew. Chem. Int. Ed.* 2017, 56, 8672–8676. [CrossRef] [PubMed]
67. Baell, J.B.; Nissink, J.W.M. Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017—Utility and Limitations. *ACS Chem. Biol.* 2018, 13, 36–44. [CrossRef] [PubMed]

68. Chakravorty, S.J.; Chan, J.; Greenwood, M.N.; Popa-Burke, I.; Remlinger, K.S.; Pickett, S.D.; Green, D.V.S.; Fillmore, M.C.; Dean, T.W.; Luengo, J.I.; et al. Nuisance Compounds, PAINS Filters, and Dark Chemical Matter in the GSK HTS Collection. *SLAS Discov. Adv. Sci. Drug Discov.* **2018**, *23*, 532–545. [CrossRef] [PubMed]

69. Jasial, S.; Hu, Y.; Bajorath, J. How Frequently Are Pan-Assay Interference Compounds Active? Large-Scale Analysis of Screening Data Reveals Diverse Activity Profiles, Low Global Hit Frequency, and Many Consistently Inactive Compounds. *J. Med. Chem.* **2017**, *60*, 3879–3886. [CrossRef] [PubMed]

70. Baell, J.B.; Holloway, G.A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* **2010**, *53*, 2719–2740. [CrossRef] [PubMed]

71. Gul, N.; Karlsson, J.; Tängemo, C.; Linsefors, S.; Tuyizere, S.; Perkins, R.; Ala, C.; Zou, Z.; Larsson, E.; Bergö, M.O.; et al. The MTH1 inhibitor TH588 is a microtubule-modulating agent that eliminates cancer cells by activating the mitotic surveillance pathway. *Sci. Rep.* **2019**, *9*, 14667. [CrossRef] [PubMed]

72. Kowarz, E.; Löscher, D.; Marschalek, R. Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines. *Biotechnol. J.* **2015**, *10*, 647–653. [CrossRef] [PubMed]

73. Izsvak, Z. Efficient stable gene transfer into human cells by the Sleeping Beauty transposon vectors. *Methods* **2009**, *49*, 287–297. [CrossRef] [PubMed]

74. Wiggers, C.R.M.; Govers, A.M.A.P.; Lelieveld, D.; Egan, D.A.; Zwaan, C.M.; Sonneveld, E.; Coffer, P.J.; Bartels, M. Epigenetic drug screen identifies the histone deacetylase inhibitor NSC3852 as a potential novel drug for the treatment of pediatric acute myeloid leukemia. *Pediatric Blood Cancer* **2019**, *66*, e27785. [CrossRef] [PubMed]

75. Dilshara, M.G.; Jayasooriya, R.G.P.T.; Karunarathne, W.A.H.M.; Choi, Y.H.; Kim, G.-Y. Camptothecin induces mitotic arrest through Mad2-Cdc20 complex by activating the JNK-mediated Sp1 pathway. *Food Chem. Toxicol.* **2019**, *127*, 143–155. [CrossRef] [PubMed]

76. Lüpertz, R.; Wätjen, W.; Kahl, R.; Chovolou, Y. Dose- and time-dependent effects of doxorubicin on cytotoxicity, cell cycle and apoptotic cell death in human colon cancer cells. *Toxicology* **2010**, *271*, 115–121. [CrossRef] [PubMed]

77. Hatcher, J.M.; Wu, G.; Zeng, C.; Zhu, J.; Meng, F.; Patel, S.; Wang, W.; Ficarro, S.B.; Leggett, A.L.; Powell, C.E.; et al. SRPKIN-1: A Covalent SRPK1/2 Inhibitor that Potently Converts VEGF from Pro-angiogenic to Anti-angiogenic Isoform. *Cell Chem. Biol.* **2018**, *25*, 460–470.e6. [CrossRef] [PubMed]

78. Khan, S.N.; Lal, S.K.; Kumar, P.; Khan, A.U. Effect of mitoxantrone on proliferation dynamics and cell-cycle progression. *Biosci. Rep.* **2010**, *30*, 375–381. [CrossRef] [PubMed]

79. Bernhart, E.; Stuendl, N.; Kaltenegger, H.; Windpassinger, C.; Donohue, N.; Leithner, A.; Lohberger, B. Histone deacetylase inhibitors vorinostat and panobinostat induce G1 cell cycle arrest and apoptosis in multidrug resistant sarcoma cell lines. *Oncotarget* **2017**, *8*, 77254–77267. [CrossRef]

80. Marshall, C.B.; Pippin, J.W.; Krofft, R.D.; Shankland, S.J. Puromycin aminonucleoside induces oxidant-dependent DNA damage in podocytes in vitro and in vivo. *Kidney Int.* **2006**, *70*, 1962–1973. [CrossRef]

81. Bruno, S.; Ardelt, B.; Skierski, J.S.; Traganos, F.; Darzynkiewicz, Z. Different Effects of Staurosporine, an Inhibitor of Protein Kinases, on the Cell Cycle and Chromatin Structure of Normal and Leukemic Lymphocytes. *Cancer Res.* **1992**, *52*, 470–473.

82. Murai, A.; Ebara, S.; Sasaki, S.; Ohashi, T.; Miyazaki, T.; Nomura, T.; Araki, S. Synergistic apoptotic effects in cancer cells by the combination of CLK and Bcl-2 family inhibitors. *PLoS ONE* **2020**, *15*, e0240718. [CrossRef] [PubMed]

Appendix B: High-Content Live-Cell Multiplex Screen for Chemogenomic Compound Annotation Based on Nuclear Morphology
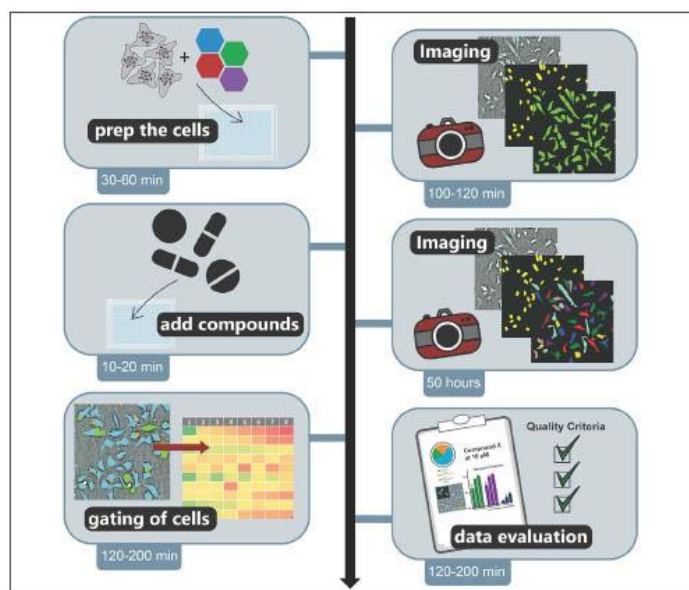
Contribution Menge, A.: Conceptualization, Methodology, Investigation, Validation and Formal Analysis, writing – Original Draft, writing, review and editing.

STAR Protocols

Protocol

High-content live-cell multiplex screen for chemogenomic compound annotation based on nuclear morphology

Amelie Tjaden, Robert T. Giessmann, Stefan Knapp, Martin Schröder, Susanne Müller

m.schroeder@pharmchem.uni-frankfurt.de (M.S.)
susanne.mueller-knapp@bmls.de (S.M.)

Highlights

A fast and flexible multiplex assay for compound annotation

Evaluate cell properties in live-cell mode

Machine learning techniques to optimize high-content data evaluation

Easily adaptable and scalable for different phenotypic features

Well-characterized small molecules enable the study of cell processes and facilitate target validation. Here, we describe a high-content multiplex screen to investigate cell viability over 48 h, which can be combined with investigating phenotypic features, such as tubulin binding and mitochondrial content, as initial cellular quality control of diverse compounds. The protocol is on a live-cell basis and easily adaptable and scalable. It details cell preparation, compound handling, plate layout configuration, image acquisition with the CQ1, and data analysis using the CellPathfinder software.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

STAR Protocols

Protocol

High-content live-cell multiplex screen for chemogenomic compound annotation based on nuclear morphology

Amelie Tjaden,[1,2,5] Robert T. Giessmann,[3,4] Stefan Knapp,[1,2] Martin Schröder,[1,2,*] and Susanne Müller[1,2,6,*]

[1]Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Max-von-Laue-Str.9, 60438 Frankfurt, Germany
[2]Structural Genomics Consortium, BMLS, Goethe University Frankfurt, Max-von-Laue-Str. 15, 60438 Frankfurt, Germany
[3]Bayer AG, Research & Development, Pharmaceuticals, 13353 Berlin, Germany
[4]Institute for Globally Distributed Open Research and Education (IGDORE), Berlin, Germany
[5]Technical contact: tjaden@pharmchem.uni-frankfurt.de
[6]Lead contact
*Correspondence: m.schroeder@pharmchem.uni-frankfurt.de (M.S.), susanne.mueller-knapp@bmls.de (S.M.)
https://doi.org/10.1016/j.xpro.2022.101791

SUMMARY

Well-characterized small molecules enable the study of cell processes and facilitate target validation. Here, we describe a high-content multiplex screen to investigate cell viability over 48 h, which can be combined with investigating phenotypic features, such as tubulin binding and mitochondrial content, as initial cellular quality control of diverse compounds. The protocol is on a live-cell basis and easily adaptable and scalable. It details cell preparation, compound handling, plate layout configuration, image acquisition with the CQ1, and data analysis using the CellPathfinder software.

For complete details on the use and execution of this protocol, please refer to Tjaden et al. (2022).

BEFORE YOU BEGIN

Here, we describe a high content multiplex screen, able to characterize generic effects regarding basic cell properties and cell viability in a multi-dimensional way, using phenotypic characterization to annotate compounds on their suitability for further phenotypic and mechanistic screening.

The present protocol describes the workflow for 135 test compounds, screened at two concentrations (1 and 10 μM) in addition to 6 control compounds in a single 384 well plate, measured for a total of 48 h at 4 different time points.

The protocol below describes the specific steps for testing osteosarcoma cells (U-2 OS). However, we have adapted this protocol to other cell lines including human embryonic kidney cells (HEK293T) and human fibroblasts (MRC-9) (see problem 1).

Culture cell lines

⊙ Timing: 2 weeks

1. Prior to the experiment, U-2 OS cells were cultured over two weeks in modified Dulbecco's Modified Eagle's medium (DMEM) plus L-Glutamine (4 mM) and high glucose (4.5 g/L D-Glucose),

supplemented with 10% fetal bovine serum (FBS) and 1% Penicillin (100 U/mL)/Streptomycin (100 μg/mL) in a corning T-75 cell culture flask.

a. The cells should be incubated at 37°C and 5% CO_2 partial pressure in a suitable incubator.

b. Constant cell growth was obtained by passaging the cells two to three times a week.

   i. The cells should be passaged when they reach 70%–80% confluence.

   ii. Cells were detached from the culture flask using trypsin (0.025% Trypsin and 0.01% EDTA), followed by quenching with the modified DMEM and diluted at a ratio of 1:30 (2–6 × 10^5 cells) in a new culture flask with fresh, modified DMEM.

c. In our experience, U-2 OS cells of passage 5–35 yield reproducible results.

Optimize cell seeding density

⏱ Timing: 48 h

The described protocol is optimized and validated for immortalized and primary adherent cell lines. Due to the autofocus mechanism of the imager, suspension cells, as well as 3D cultures, will require additional optimization steps and may not be suitable for this protocol (see problem 1). Live cell imaging is best accomplished when cells in the wells have a confluence ranging from 50 to 90%. High confluence can result in difficulties in the segmentation of individual cells. Low confluence can affect the statistical power of the data (see problem 2) and the overall cell behavior. Depending on the length of the experiment, cell densities should be adapted accordingly. The starting confluence should be chosen such that robust segmentation of individual cells can still take place after 48 h i.e., cells should not be overgrown at that time point. We conducted an experiment before the start of the experiment to determine the best seeding concentration of U-2 OS cells for optimal cell segmentation and cell growth (see Figure 1). We therefore recommend that the confluence at the first time point should be more than 40% and after 48 h the confluence should not exceed 90%.

2. Optimize cell seeding density by setting up a preliminary experiment with varying cell densities and monitor their confluence over 48 h.

a. Wash the cells cultured in a T75 culture flask once with 5 mL PBS buffer.

b. Detach the cells using 3 mL trypsin and determine the cell count and viability of the cells using trypan blue 0.4% solution in a Neubauer chamber or an automated cell counter such as the TC20 Automated Cell Counter (Bio-Rad).

   i. Mix 10 μL cells and 10 μL trypan blue and add to a counting slide for automated cell counting.

   ii. Count the unstained (live) and stained (dead) cells.

   iii. Viability of the cells should be more than 95%.

c. Dilute the cells into 6 different concentrations ranging from 2,500 cells/well to 1,000 cells/well in steps of 250 cells/well.

d. Pipet 50 μL/ well of diluted cells according to the layout shown in Table 1 in a clear bottom, black, 384-well plate.

e. Create 3 mL of a stock solution with 50,000 cells/mL for a final count of 2,500 cells/well.

   i. For 2,500 cells pipet 50 μL of stock solution in row 2 B to G.

   ii. For 2,250 cells pipet 45 μL of stock solution and 5 μL of media in row 3 B to G.

   iii. For 2,000 cells pipet 40 μL of stock solution and 10 μL of media in row 4 B to G.

   iv. For 1,750 cells pipet 35 μL of stock solution and 15 μL of media in row 5 B to G.

   v. For 1,500 cells pipet 30 μL of stock solution and 20 μL of media in row 6 B to G.

   vi. For 1,250 cells pipet 25 μL of stock solution and 25 μL of media in row 7 B to G.

   vii. For 1,000 cells pipet 20 μL of stock solution and 30 μL of media in row 8 B to G.

   viii. Pipet 100 μL/well of PBS or any other buffer in the outer wells to minimize evaporation variations (see problem 3).

STAR Protocols
Protocol
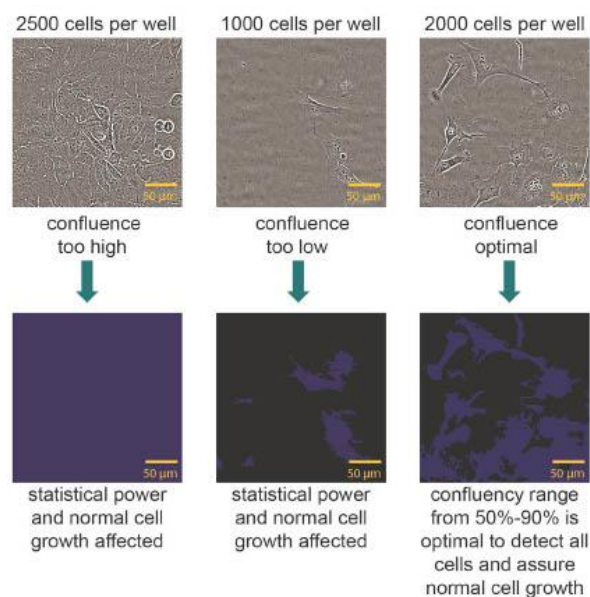
Figure 1. Confluence display of U-2 OS cells
Brightfield image of U-2 OS cells stained after 48 h using the Cellcyte X live-cell imaging system. Images show U-2 OS after different cell seeding concentrations (2500 cells/well, 1000 cells/well and 2000 cells/well) and the detection of the confluence using the implemented Cellcyte X analysis software.

3. Image the plate one day after seeding (time point 0 h) and 48 h after the first imaging. For this step any light microscope can be used. Here we used the Cellcyte X Live-cell imaging system.

   Note: The first imaging will represent the baseline time point of the later experiments. The second and third imaging step is important to validate the confluence after 48 h.

   a. Image the plate once 24 h after seeding and 48 h after time point 0 h using the brightfield channel at 10× magnification.

4. Determine confluence levels at both time points (Figure 1).

Table 1. Layout to determine optimal cell seeding density

384 plate	1	2	3	4	5	6	7	8
A	Buffer	buffer	buffer	buffer	buffer	buffer	buffer	buffer
B	Buffer	2500	2250	2000	1750	1500	1250	1000
C	Buffer	2500	2250	2000	1750	1500	1250	1000
D	Buffer	2500	2250	2000	1750	1500	1250	1000
E	Buffer	2500	2250	2000	1750	1500	1250	1000
F	Buffer	2500	2250	2000	1750	1500	1250	1000
G	Buffer	2500	2250	2000	1750	1500	1250	1000

a. Either use the software implemented in the microscope used or any other software that can detect the confluence level of the cells using the brightfield channel. We used the implemented Cellcyte X analysis software.
   i. Train the software with 5–10 example wells.
   ii. Use wells from different columns to increase the variability of the confluence levels.
   iii. Analyze the whole plate and check by eye on 5–10 other wells, to see if the mask detected all cells.
b. The confluence at the first time point should be sufficient for normal cell growth.
c. Confluence after second time point should not exceed 90%.
5. Select the best concentration of cells, at which you can see normal cell growth (Musa et al., 2013) and the confluence after 48 h does not exceed 90%.

⚠ CRITICAL: Ensure that the cells are seeded at an appropriate density and are evenly distributed within the wells to avoid malfunction of the autofocus and ensure statistical evaluability (see problem 6). To make sure the cells are evenly distributed when attaching to the plate, leave the plate at room temperature for 30 min before placing it in the incubator. For best measurement results, the plate should be left in the incubator for a few minutes before being imaged.

Preparation of compounds and plate layout

🕐 Timing: 5 min per compound

The compounds of interest as well as specific reference compounds necessary for the experiment can be prepared in advance. If using the Echo 550 automated liquid handler, we recommend preparing a source plate and a "pick list". The reference compounds are necessary to train the machine learning algorithm. For each cell property, at least one reference compound must be tested. As part of the assay development, we tested the described reference compounds in U-2 OS cells (Tjaden et al., 2022) and confirmed their suitability to generate a training dataset (see data analysis CellPathfinder software).

6. All compounds should be dissolved and/or diluted in a suitable solvent such as DMSO to a 1000-fold higher concentration than the final assay concentration. This ensures a minimal final concentration of 0.1% DMSO which has been found to produce no measurable effect (Brayton 1986) on the proliferation of U-2 OS cells.

Note: If other cell lines are used, the DMSO sensitivity should be tested beforehand (see problem 4).

7. If using an automated acoustic liquid handler, such as an Echo 550 (LabCyte) the compounds should be added to a compatible source plate, e.g., 384-well LDV Source plate (Labcyte).
   a. Add a minimum 3 μL of 1000× compound solution into each well of the source plate.
   b. Centrifuge the LDV source plate once at 500 × g for 3 min.
   c. To minimize compound contamination, the plates should be sealed with DMSO resistance seals (VWR) after each usage.
   d. The plates can be stored at -20°C over a longer time period (for at least 2 years on average).
      i. One compound plate can be used multiple times, when stored correctly.
      ii. Before reusing the plate, ensure that the compounds are completely thawed. We recommend leaving the compound plate at least one hour at room temperature before pipetting. Also, precipitation of compounds should be checked by eye, because precipitation of compounds within the source plate can lead to false negative results (see problem 4).
      iii. Check also if there is enough compound solution left for your experiment. You can check that using the survey detection function included in the Echo550 Labcyte software.

STAR Protocols
Protocol

Table 2. Reference compounds with mode of actions and gating steps they are used for

Reference compound	Mode of action	Used for what gating step
staurosporine	promiscuous kinase inhibitor (Bruno et al., 1992)	apoptotic cells, pyknosed nuclei
paclitaxel	targets tubulin/no disassembly of mitotic spindle (Wang et al. 2000)	Change of tubulin structure
milciclib	CDK inhibitor (Sanchez-Martinez et al., 2015)	Increase of mitochondrial mass
daunorubicin	anthracycline antibiotic, intercalation of DNA strands, ROS production (Al-Aamri et al., 2019)	Increase of mitochondrial mass, fragmented nuclei, apoptotic cells
digitonin	Detergent (Styrt et al. 1985)	Permeabilization of plasma membranes
berzosertib	ATR/ATM-inhibitor (Fokas et al., 2012)	Hoechst high intensity objects

iv. To minimize absorption of the compound into the wall of the storage plate, use polypropylene plates or other plates recommended for compound storage.

8. For the analysis, use the following reference compounds at a final concentration of 10 μM: staurosporine, paclitaxel, milciclib, daunorubicin, digitonin and berzosertib (see Table 2).
   a. Dissolve reference compounds in DMSO to a final concentration of 10 mM.
   b. Add a minimum 3 μL of each of these reference compounds to an individual well of your LDV source plate when pipetting with the Echo 550.
9. Prepare a pick list by using the "Pick list" program for the Echo 550 according to the manufacture's manual (manual can be found here: https://www.agilent.com/cs/library/usermanuals/Public/G5415-90026_DDUG_EN.pdf).
   a. A possible layout of the pick list is shown in Table 3.
10. For an example of a pick list refer to Table S1 (Zenodo: https://zenodo.org/record/7092795#.YyhponZByUk).
11. Add the compounds as well as the reference compounds to your pick list so that they will be added to your destination plate.
   a. The reference compounds should be added in technical quadruplicates (see Figure 2) to get a sufficient validation set for the analysis protocol (see data analysis CellPathfinder software).
   b. Compounds should be pipetted to the destination plate at the concentration of interest in a final volume of 50 μL. The standard protocol assesses compounds at 10 μM and 1 μM. Thus, 50 nL and 5 nL of inhibitor stock solution should be transferred, respectively (see example in Table S1 (Zenodo: https://zenodo.org/record/7092795#.YyhponZByUk)). Due to the large dilution factor, the volume of 45 nL DMSO can be neglected and a final concentration of 0.1 vol% of DMSO is not affecting cell (Brayton 1986) proliferation.
   c. Compounds are tested in technical duplicates.
   d. 5%–10% of the plate should be "blank" wells filled with DMSO 0.1% i.e., by transferring 50 nL of DMSO per well. This is similar to the dilution factor for compounds from 10 mM -> 10 μM.
12. A possible layout of the plate can be found in Figure 2.
13. Save your picklist as ".csv" on your local storage space.

⚠ CRITICAL: Compounds with limited solubility in DMSO can precipitate in the source wells and will therefore not be dispensed to the cells. Pay attention to minimize freeze-thaw-cycles, because this can affect the stability of the compound solutions. Test DMSO solubility of the compound stock solutions before testing the compound. If not soluble in DMSO, try diluting in water and pipet by hand or with ECHO LabCyte protocol for aqueous solvents (384LDV_aq). (See problem 4).

Table 3. Layout Pick list for Echo550 Liquid Handling System

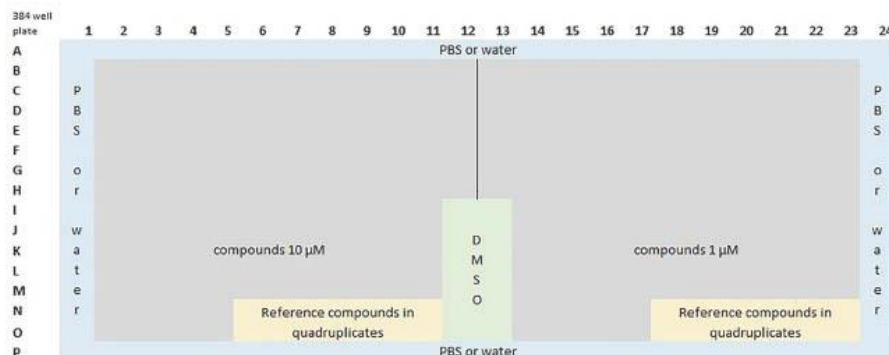| Source plate name	Source plate well	Destination plate name	Destination plate well	Transfer volume (nL)	Compound description
Source Plate 1	A01	Destination Plate 1	B02	50	Compound 1
Source Plate 1	A02	Destination Plate 1	B03	50	Compound 2
Source Plate 1	A03	Destination Plate 1	B04	50	Compound 3
Source Plate 1	A04	Destination Plate 1	B05	50	Compound 4

Figure 2. Possible Layout of one 384-well plate to test 135 compounds at two different concentrations

Determination of channels and scan parameters for imaging protocol

🕐 Timing: 1 h

For image acquisition, the protocol for the CQ1 can be found under the name "Measurement protocol for the Multiplex assay using the CQ1 and files containing the settings to operate the CQ1 microscope with Momentum" at Zenodo: https://doi.org/10.5281/zenodo.6394521. However, for every change in the assay protocol (different plate, different cell line, different conditions etc.) or when using a different microscope, the protocol should be adjusted accordingly.

14. Here are the standard settings in detail:
    a. Magnification: A 10× objective is used, as it will capture a large surface area while retaining resolution to capture cell morphology. Higher magnifications like 20× can also be used to obtain a more detailed image, but with the consideration that fewer cells can be imaged. Lower magnification of 2× might not detect detailed phenotypic properties.
    b. Five channels are used for image acquisition: Ex 405 nm/Em 447/60 nm, 500 ms, 50%; Ex 561 nm/Em 617/73 nm, 100 ms, 40%; Ex 488/Em 525/50 nm, 50 ms, 40%; Ex 640 nm/Em 685/40, 50 ms, 20%; brightfield, 300 ms, 100% transmission.
    c. Focus and Z-stacks: The area where the Z-stacks will be set can vary from plate to plate. Use 7 Z-stacks for optimal cell imaging with a total of 55 μm spacing to ensure the compensation of potential plate variations and generation of a robust readout. If using the autofocus for your experiment, it may happen that the microscope will sometimes focus on dust and will lose focus (see problem 6), so we recommend to set the focus area for 384-well plates with a height of 14,4 ± 0,1 mm at -15.5 μm for the lowest (-20.2 μm) and highest focus of 39.7 μm (51.6 μm), respectively.

Note: All parameters should be adjusted when details in the set-up are changed e.g., when other cells, plates or other dyes are used. Also, the time frame can freely be adapted depending on the experiment. For viability studies an incubation time of at least 24 h is recommended.

⚠ CRITICAL: Use the same plates that are implemented in the CQ1 measuring software. If the plate differs, the focus area might not be as accurate as intended (see problem 6).

STAR Protocols

Protocol

The specific plate parameters must be implemented in the CQ1 software (refer to the manufacturer's manual, which can be found here: https://www.ibios.uni-osnabrueck.de/index.php?cat=Light%20Microscopy/Yokogawa%20CQ1&file=CQ1_User_Manual.pdf).

Preparation of the data analysis workflow

⏲ Timing: 2 h

15. Install Python 3 on your computer, preferentially via the distribution Anaconda.
    a. Anaconda can be downloaded from https://www.anaconda.com.
    b. Follow the Getting started instructions at https://docs.anaconda.com/anaconda/user-guide/getting-started/ to open the Anaconda Prompt.
    c. Note the path displayed in the Prompt. This will be named "root path" in this protocol.
16. To set up your Python environment and prepare the analysis scripts, visit Zenodo: https://doi.org/10.5281/zenodo.6325622 and follow the instructions given.

Note: The exact version and specific DOI of the software you will be using by checking the "Versions" box in Zenodo, and cite the Zenodo record correspondingly in your work (e.g., v1.0.1 of the scripts has the DOI: Zenodo: https://doi.org/10.5281/zenodo.6327204).

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
staurosporin	Selleckchem	Cat#: S1421
paclitaxel	Selleckchem	Cat#: S1150
digitonin	Invitrogen	Cat#: BN2006
milciclib	Selleckchem	Cat#: S2751
daunorubicin	Selleckchem	Cat#: S3035
berzosertib	Selleckchem	Cat#: S7102
Hoechst33342	Thermo Scientific	Cat#: 62249 https://www.thermofisher.com/order/catalog/product/62249
BioTracker™ 488 Green Microtubule Cytoskeleton Dye	EMD Millipore	Cat#: SCT142 https://www.merckmillipore.com/DE/de/product/BioTracker-488-Green-Microtubule-Cytoskeleton-Dye,MM_NF-SCT142
MitoTracker red	Invitrogen	Cat#: M22425 https://www.thermofisher.com/order/catalog/product/M22425
Annexin V Alexa Fluor 680 conjugate	Invitrogen	Cat#: A35109 https://www.thermofisher.com/order/catalog/product/A35109
Trypan blue 0.4%	Thermo Fisher Scientific	Cat#: 15250061 https://www.thermofisher.com/order/catalog/product/15250061
Critical commercial assays		
alamarBlue™ Cell Viability Reagent	Invitrogen	Cat#: DAL1025
Deposited data		
raw images and processed images	BioImage Archive	S-BIAD145
protocol establishment information	Tjaden et al.	https://doi.org/10.20944
Experimental models: Cell lines		
U-2 OS cells	ATCC	HTB-96™
Software and algorithms		
CQ1 microscope software	Yokogawa	v 1.04.03.01
CellPathfinder Software v3.04.02.02	Yokogawa	N/A
Python v3.9	Python Software Foundation	N/A

(Continued on next page)

CellPress
OPEN ACCESS

STAR Protocols
Protocol

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Microsoft Excel v16.0.4266.1001	Microsoft	N/A
GraphPad Prism v8.4.3	GraphPad Software	N/A
Cellcyte Studio	CYTENA	v2.6.0
Scripts for data analysis	Python	Zenodo: https://doi.org/10.5281/zenodo.6325622
CellPathfinder analysis protocols	Yokogawa	Zenodo: https://doi.org/10.5281/zenodo.6415330
Other		
CQ1 microscope	Yokogawa	N/A
384-well cell culture microplate, PS, f-bottom, µClear®	Greiner	Cat#: 781091
ECHO® 550 Acoustic Liquid Handler	Labcyte	N/A
ECHO® source plate	Labcyte	Cat#: P.05525
Cytomat2C24 incubator	Thermo Scientific	N/A
TC20 Automated Cell Counter	Bio-Rad	https://www.bio-rad.com/de-de/product/tc20-automated-cell-counter?ID=M7FBG34VY
CELLCYTE X™	CYTENA	v6.1.1
DMSO resistance seals	VWR	391-0642
Supplemental Material		Zenodo: https://zenodo.org/record/7092795#.YyhponZByUk

MATERIALS AND EQUIPMENT

This protocol should be adaptable to other confocal microscopes, which have similar imaging capacities than the Yokogawa CQ1. However, the image acquisition and analysis would need to be adapted accordingly.

For data analysis, other software, such as the open-source software "deepImageJ" with an implemented deep learning function (https://deepimagej.github.io/deepimagej/) or other built-in software can be used, though the described analysis will no longer be applicable and further data reformatting steps might be necessary.

> *Optional:* As an alternative to a liquid handling system, manual pipetting was tested. The volume and concentration of the chemicals and compounds have to be adapted accordingly to transfer volumes accurately.

The CQ1 can also be implemented in an automation system like Momentum v5.3.

This protocol has been validated against three different adherent cell lines, immortalized and non-immortalized (Tjaden et al., 2022).

STEP-BY-STEP METHOD DETAILS
Addition of fluorescent dyes and cell seeding

> ⏱ Timing: 45–60 min

These steps describe the preparation of cells, the seeding and the following addition of the fluorescent dyes to the cells.

1. Dilution of U-2 OS cells for seeding of one 384 well plate.
   a. Wash the cells in a T75 flask once with 5 mL PBS buffer.
   b. Detach the cells using 3 mL trypsin and determine the concentration and viability of the cells using trypan blue 0.4% solution.
      i. Mix 10 µL cells and 10 µL trypan blue and add to a counting slide for automated cell counting.
      ii. Put the counting slide into the TC20 Automated Cell Counter.

iii. Count the unstained (live) and stained (dead) cells.
iv. Wait till the results appear on the Cell Counter screen.
v. Viability of the cells should be higher than 95%.
c. Dilute the cells to 3 × 10⁴ cells/mL and a final volume of 16 mL/384 well plate or the cell concentration determined in your preliminary experiment for a confluence ranging from 50%–90% after 48 h.

Note: For a 384-well plate format and 50 μL per well, the cell concentration range is typically between 2–4 × 10⁴ cells/mL (1,000–2,000 cells/well). Depending on the cell type and experimental setup this can be adapted using the steps described above see optimize cell seeding density.

2. Addition of the fluorescent dyes to the cells.
   a. Add to the 16 mL cell suspension prepared in step 1c:
      i. 57 μL of a 16.23 μM stock of Hoechst 33342 (final assay concentration of 60 nM).
      ii. 8 μL of a stock of BioTracker™ 488 Green Microtubule Cytoskeleton Dye.
      iii. 12 μL of a 100μM stock of Mitotracker red (final assay concentration of 75 nM).
      iv. 93 μL of a stock of Annexin V Alexa Fluor 680 conjugate.

Note: Using different fluorescent dyes than the ones mentioned here, can potentially lead to bleed-through problems. For the here described protocol, the overlapping of fluorescence emission spectra is neglectable for all dyes but the MitoTracker Red and Annexin V Alexa Fluor 680. However, this overlap does not influence the analysis, since the excitation maxima of these two dyes are well separated and the gating. See also (Tjaden et al., 2022).

   b. Mix gently by inverting the cell suspension.
   c. Transfer 50 μL of the cell suspension with fluorescent dyes prepared in step 2 to the corresponding wells of the 384 well plate using a multichannel pipet and a reagent reservoir. The plate layout can be found in Figure 2.
   d. Pipett 100 μL PBS buffer or water in the outer wells as indicated in Figure 2.

Note: This step helps avoiding edge effects due to evaporation (see problem 3).

3. Remove any air bubbles that may have been formed during pipetting (see Figure 3). This step can be performed using an ethanol bottle where a straw-like nozzle is positioned above the ethanol surface within the bottle. Gently press air on the surface of the cell dilution in the wells to remove the air bubbles.
4. Put a lid on the plate prepared in steps 3–6 and leave it for 30 min under the cell culture hood at room temperature to allow reattachment of the U-2 OS cells to the plate.
5. Incubate the plate for 18–24 h at 37°C and 5% CO₂.

Figure 3 **Debubbler.** Example illustration of air bubble removal.

⚠ CRITICAL: Ensure an even and reproducible distribution of cells within the wells. Image acquisition should ideally be performed in the center of the well to ensure the best coverage of the well area. Uneven distribution of cells within a well can at a later point impede the statistical analysis (see problem 5).

⏸ Pause point: The following steps will be conducted the next day.

Image acquisition of non-treated cells

⏱ Timing: 3–4 h

wells containing air bubbles

wells containing no air bubbles

Ethanol
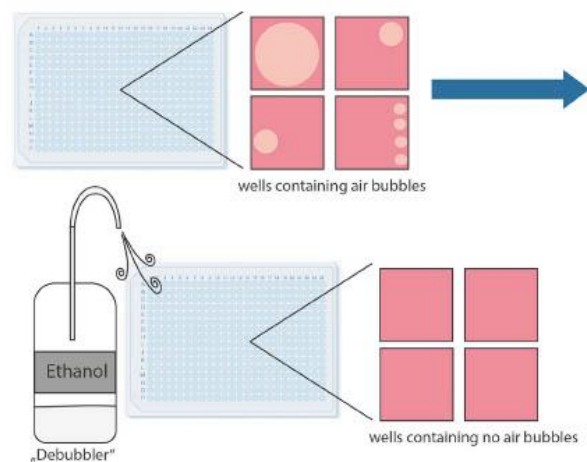
„Debubbler"

Figure 3. Debubbler
Example illustration of air bubble removal.

This step describes the image acquisition of the non-treated cells. The resulting data serve as a baseline for the statistical normalizations (see evaluate your data).

6. Switch on the CQ1 microscope and the separate laser module.
   a. Wait 3 min for the laser to heat up.
7. Ensure that the environmental controls regulating temperature, CO_2 partial pressure as well as the water reservoirs in the CQ1 microscope and the plate incubator are functioning.
   a. The temperature of the CQ1 microscope should be kept stable at 37°C.
   b. The partial pressure of CO_2 should be set to 5%.
   c. The humidity should be set to 100% to avoid evaporation (see problem 3).
8. Place the plate prepared in steps 3–6 in the CQ1 and open the imaging software using either the associated automation unit or manual handling.
9. Open the measuring protocol, described in the section determination of channels and scan parameters for imaging protocol.
10. Assess if the measuring protocol provides focused images of sufficient intensity in all channels.
    a. Choose randomly approximately 10 wells spread over the entire plate.
    b. Test if the focus area and channel parameters can detect the cells and fluorescent signals of the different channels (see problem 6).
11. Measure the plate once with non-treated cells using the described protocol.
12. Put the plate into an incubator at 37°C and 5% CO_2 until compound treatment.

   Note: The time between the measurement and the compound treatment should be as short as possible.

Compound treatment

   © Timing: 30–60 min

STAR Protocols
Protocol

This step describes compound treatment of the cells.

13. Use the Echo® source plate (see preparation of compounds and plate layout) prepared in step 7 to add the compounds to your plate.
14. Use the "Pick list" program in the Echo® 550 liquid handler system to aliquot the correct volumes of the inhibitors and DMSO into each reaction wells in the assay plate (see 9–13 in preparation of compounds and plate layout).
    a. Open the "Pick list" program for the Echo 550 system on your computer.
    b. Select "new" to create a new protocol.
    c. Enter for the source- and destination plate the right Sample Plate Format (384LDV) and the right Sample Plate Type.
       i. For DMSO as a solvent use 384LDV_DMSO as plate type.
    d. Now you can import your picklist, which should be saved as ".csv" on your local storage space.
       i. Check your layout again on "plate preview".
15. After adding the compounds, centrifuge the plate for 3 min at 100 × g.

Image acquisition treated cells

⏱ Timing: 24–48 h

This step describes the image acquisition of the treated cells over a defined time.

16. Make sure that the CQ1 microscope is prepared as described in steps 8 and 9 of Image Acquisition of non-treated cells and place the plate in the CQ1.
17. Open the measuring protocol used for the acquisition of the non-treated cells (see image acquisition of non-treated cells).

    Note: For image acquisition of treated cells, the same measuring protocol should be used as is for non-treated cells. It is important that the same cell area will be imaged and therefore robust baseline values can be calculated.

18. Test the measuring protocol with the inserted cell plate.
    a. Therefore, scan 3–5 wells to test if the focus area of the initial baseline measurement is still appropriate (see problem 6). If the CQ1 cannot detect the cells, the plate might not be placed properly into the machine or there was a problem while pipetting (e.g., air bubbles occurred).
    b. Define Scan Schedule and Frequency of your testing. Here, we measure the plate 12 h, 24 h and 48 h after compound treatment (see Figure 4).
19. To free instrument time of the CQ1 microscope, the plate can be also transferred to an incubator (37°C, 5% CO_2).

Data analysis CellPathfinder software

⏱ Timing: 4–5 h

This step describes the image analysis using the CellPathfinder software (Yokogawa) and is specific to the CQ1 microscope. Similar high-content microscopes with their own built-in software for analysis or open-source analysis software should be compatible with this protocol, however will require hard- and software related adaptation. For more comprehensive instructions regarding the Pathfinder software, consult the Yokogawa's product specialists and technical documents. The resulting data and processed images will be evaluated afterwards (see evaluate your data) to obtain a phenotypical annotation of each tested compound. The protocol we used to analyze U-2 OS cells
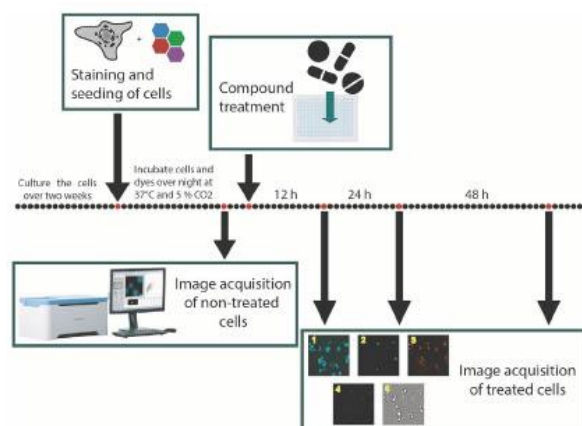
Figure 4. General time schedule
Schemed time schedule of one plate, which is measured before compound treatment and 12 h, 24 h and 48 h after compound treatment using the CQ1 confocal microscope.

"CellPathfinder_AnalysisProtocol_U2OS_Multiplex" can be found here: Zenodo: https://doi.org/10.5281/zenodo.6415330.

> Note: The generated protocol is specific to the cell type and magnification. Experiments with different components will require generating a new protocol.

20. Open the CellPathfinder software (see Figure 5A).
21. In the Data tab you can see the measurement details of your experiment.
22. Open the data you want to analyze first.
    a. Open the folder on your computer where you saved the CQ1 measurement data.
    b. Select the data that you want to analyze first.
    c. Convert the CQ1 image data by clicking "convert".
23. It is recommended to create an analysis for each cell line individually.
24. Open the Protocol tab and click on "Load" and select the analysis protocol for U-2 OS cells "CellPathfinder_AnalysisProtocol_U2OS_Multiplex".Now you can go step by step through the parameters by clicking on the process buttons at the top of the operation window (see Figure 5A).
25. Open the "Channels" section. We are using 6 different channels (CH) (see Figure 5B), based on our study design. The brightfield is used twice.
    a. CH1: Hoechst33342 (DNA detection): Ex 405 nm/Em 447/60 nm.
    b. CH2: BioTracker™ 488 Green Microtubule Cytoskeleton Dye (tubulin stain): Ex 488/Em 525/50 nm.
    c. CH3: MitoTracker red (mitochondrial mass detection): Ex 561 nm/Em 617/73 nm.
    d. CH4: Annexin V (apoptosis marker): Ex 640 nm/Em 685/40.
    e. CH5: brightfield, DCP Fluor Type Simple Mode, Contrast 0.06.
    f. CH6: brightfield, DCP Phase Type Simple Mode, Contrast 0.04.
26. Channels 5 and 6 are set on DCP to create the best digital phase contrast (see Figure 5C).
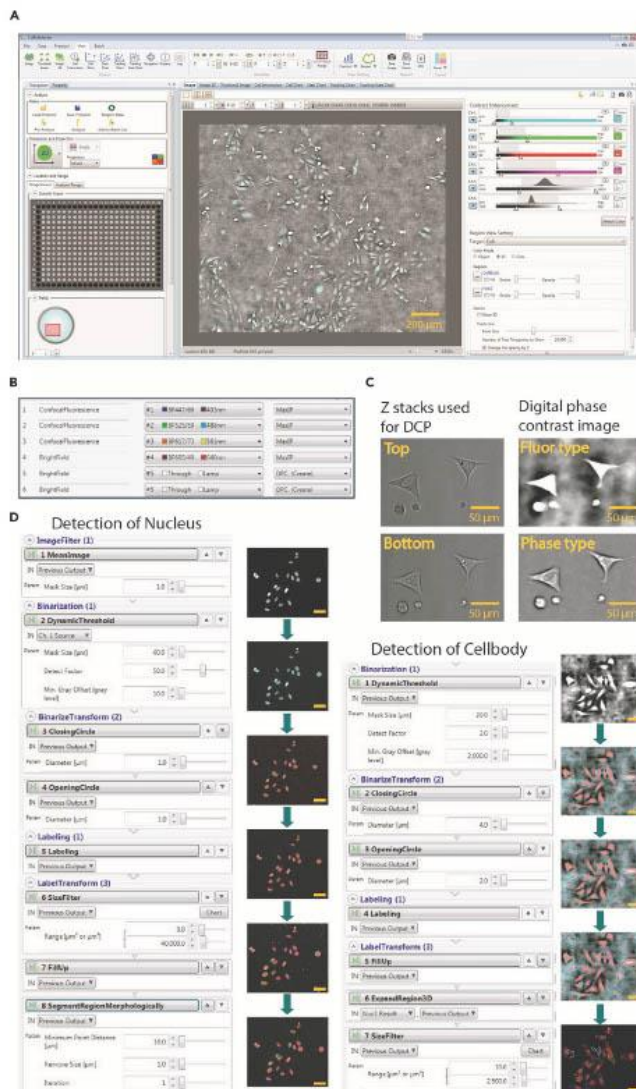
**Figure 5. Protocol generation in CellPathfinder software**
(A) Operation window for CellPathfinder software.
(B) Channels used for analysis.
(C) Z-stacks used for digital phase contrast (DCP) and DCP as fluo- and phase type.
(D) Algorithm used to detect nucleus and cell body, as well as example images.

Note: The "fluortype" setting of the digital phase contrast (DPC) in CH5 provides better segmentation of the cells against the background while the phase setting in CH6 can be used to generate high contrast cell images for visual inspection or publication figures.

 a. For both digital phase contrast channels, two Z stacks should be used to be overlaid and get the best contrast image to detect cellular shape. Here, we used Top Z 5 and Bottom Z 3. Depending on the plate, these Z stacks might differ.

27. For Object assembly open "Object". Here we created 2 different objects ("Nucleus" and "Cell body").
 a. "Nucleus": Ex 405 nm/Em 447/60 nm, Finder: Nuclear, Recognition: Advance.
 b. "Cell body": brightfield, Finder: Cell, Recognition: Advance.
28. A process based on the defined objects was created (Figure 5D). The aim of the algorithm is to detect cellular shape "Cell body", based on the brightfield and cell nuclei "Nucleus" based on fluorescent signal of the Hoechst channel. For the Cell body we used the brightfield channel with the fluor-type (CH5).
29. The Nuclei and the Cells are linked in the "Link" section.
 a. Included nucleus in cell body.
 b. One in one (1:1).
30. To proceed with the protocol in the CellPathfinder software, you have to first pre-analyze your data. This step includes image processing and object detection and is inevitable to later use the machine learning-based functions for the cellular gating. For an easier and faster workflow, you can pre-analyze just wells with reference compounds important for training of the machine learning (ML)-based functions as well as non-treated cells (see preparation of compounds and plate layout).
 a. The pre-analyzing of important wells should secure the detection of cell bodies as well as cell nuclei. Check if the cell bodies and nuclei are correctly detected in 5–10 example wells.
 b. Click on "Pre-Analyze" and select the wells including the reference compounds (N5 to N12 and O5 to O12) (see preparation of compounds and plate layout).
31. After pre-analysis, the ML based gating algorithm can be trained by opening the "Gate" section.
32. All features we used can be found in the Table S2 (Zenodo: https://zenodo.org/record/7092795#.YyhponZByUk).
33. The machine learning based workflow of the analysis is shown in Figure 6.

Note: To detect the phenotypic changes based on healthy nuclei, a tree principle is used. For a detailed explanation on the gating principle and all features used, see (Tjaden et al., 2022).

 a. The cells will be gated first into the categories Hoechst High Intensity Objects "HighIntObj" or normal cells "Normal".
 b. All cells that are gated "Normal" will then be gated into cells containing a healthy nucleus "HealthyNuc", a pyknosed nucleus "PyknoNuc" or a fragmented nucleus "FragNuc".
 c. A pyknosed nucleus can either appear due to mitosis or apoptosis. Thus, cells with pyknosed nuclei will be gated further into "mitosis" or "apoptosis".
 d. To investigate the phenotypic properties of interest (tubulin effect, mitochondrial mass increased, membrane permeabilized), cells showing a healthy nucleus will be gated further in "tubulin effect" or "tubulin normal", "mito mass high" or "mito mass normal" and "membrane permeabilized" or "membrane intact".

STAR Protocols
Protocol
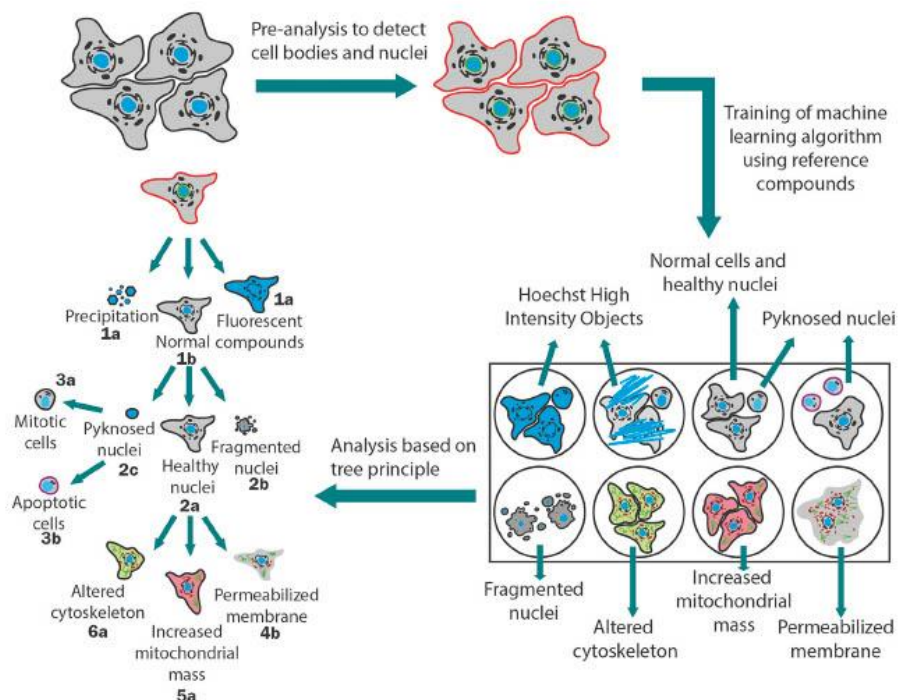
Figure 6. Workflow of machine learning based analysis using the CellPathfinder software
The cell bodies and cell nuclei are detected in the pre-analysis of the software. Afterwards the machine learning based algorithm can be trained based on the reference compounds (see preparation of compounds and plate layout). With the trained protocol, the whole plate and all time points can be analyzed based on the tree principle explained in Tjaden et al. (Tjaden et al., 2022).

34. Use the reference compounds with known phenotypes to train the machine learning -based analysis (see Table 2).
    a. For each gating step, individual cells (minimum 10 images) are selected to serve as an example for the gating phenotype (Figure 7). This can be done by a trained scientist or the example images provided here can be used.
35. After selecting the example images for the ML-based analysis, safe your analysis and run it once on all other wells containing reference compounds.
36. Ensure that the example images worked appropriately to create a robust analysis by comparing the results of all reference compound treated wells (see problem 9). The results should be consistent with the known biological mode of action of the reference compounds and should not differ from each other by more than 10%.
37. If you are sure that the machine learning-based analysis has been well trained (no deviation of the results of the reference compounds by more than 10%), the whole plate can be analyzed.
38. Your results in from of separated csv files can be found in the folder containing the data under "...\PFAnalysis" from where it can then be further evaluated (see evaluate your data).

STAR Protocols
Protocol
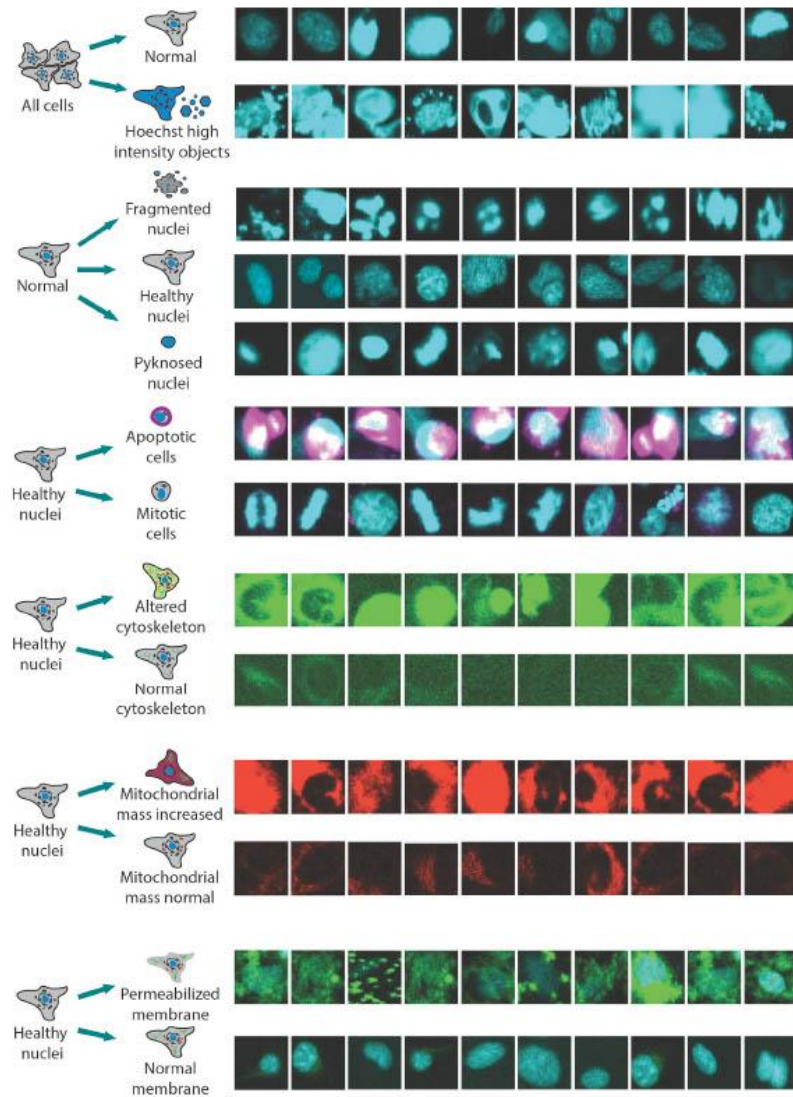
Figure 7. Machine Learning based algorithm training images
For every gating step, 10 example images of the chosen cells are shown here. The images show the most relevant fluorescent channels for the related gating step. The cells were selected by a trained scientist.

Optional: Generation of processed images

⊘ Timing: 2–3 h

This is an optional step. For publication and/or deposition of your data, we recommend generating processed images of your experiments (see Figure 8).

39. Open the "View" tab on the CellPathfinder software.
40. Use the contrast enhancement to adjust channel intensity levels for your image export.
41. Save Images as .jpg in a folder called "processed images".
42. name the files according to the experimental-ID, cell line and time point e.g., CQ1-ctf006-U2OS-0h.

Document your data

⊘ Timing: 1 h

This step describes how to document the generated data so it can be evaluated automatically in the next step. For preparation of your computer refer to preparation of the data analysis workflow.

43. Visit the "root path" (see preparation of the data analysis workflow) on your computer.
44. Open the metadata template file found here: Zenodo: https://doi.org/10.5281/zenodo.6325622 (excel file with multiple sheets) called "metadata_template.xlsx". Do not change any column headings. You are free to add additional columns.
45. Fill all sheets.
    a. In the sheet "experiments", describe your basic experimental setup.
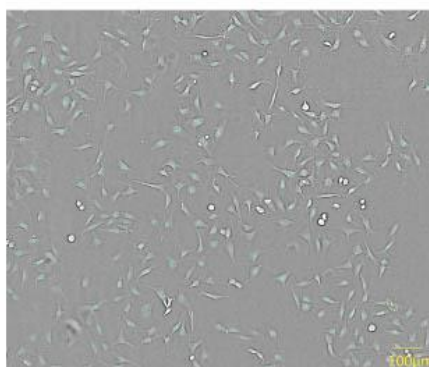
Figure 8. Processed Image of U-2 OS cells
Brightfield confocal image of stained (blue: DNA/nuclei, green: microtubule, red: mitochondria content, magenta: Annexin V apoptosis marker) U-2 OS cells after 24 h of exposure to 0.1% DMSO.

STAR Protocols
Protocol

b. In the sheet "imaging campaigns", describe all time points for which you acquired images. Make sure to fill the column "cq1 analysis available in folder"; this can be a local, absolute path on your computer (e.g., "C:/Users/CQ1/experiment/PFAnalaysis/timestamp_protocol/") or can be indicated relative to the "root path" (e.g., "../../experiment/PFAnalaysis/timestamp_protocol/").

Note: This cannot be a network path. If necessary, copy the files from a network path to your computer. See also Section "data analysis CellPathfinder software" above.

c. The sheet "exclude from file list" is only relevant if you want to deposit your data. In that case, you can omit individual processed images from the autogenerated file list.
d. In the sheet "compounds", describe all compounds you used in the experiment, e.g., with their chemical structure formula as SMILES. It is recommended to include InChls, too.
e. In the sheet "compound batches", list all specific batches of the compounds you used. Information may include internal identifiers and supplier lot numbers.

46. Each experiment is linked from the sheet "experiments" to a specific "compound map". This "compound map" describes the compound batches in individual wells, as discussed in Section "compound treatment". Example:
a. Let us assume that you entered "multiplex1" as the value of "compound map see corresponding excel table" on sheet "experiments".
b. Create (or rename) a compound map sheet to "compound map multiplex1". If you entered "multiplex2", you should name it "compound map multiplex2".
c. In the sheet "compound map multiplex1" (or alike), fill out information about the concentration of the compound and of your intent, i.e., its experimental type:
i. "blank": a well with 0.1% DMSO.
ii. "cells only": a well without any addition apart from cells.
iii. "chemogenomic candidate": a well with a compound to be tested.
iv. "control": a well with a reference compound.

Optional: If you want to deposit your data, fill out "raw data available in zip file" and "processed images available in folder", but relative to the folders in BioImage Archive (see also Section "optional: deposit your data" below).

47. Save the file under a name which will be denoted "METADATA_FILENAME" below.
48. Perform automated checks on the file to make sure that it is intact. Open the Anaconda Prompt and enter:

```
> python quality_control_of_excel_file.py METADATA_FILENAME
```

49. Make sure to replace "METADATA_FILENAME" with the corresponding filename. If you saved the file under the name "overview.xlsx", the correct command is:

```
> python quality_control_of_excel_file.py overview.xlsx
```

50. If you receive error messages, correct your file accordingly. If the error message appears to be cryptic to you, contact the authors of this protocol. If you do not see any error messages, your file is ready for the next steps.

Optional: Deposit your data

🕑 Timing: Multiple days, depending on your upload rate

STAR Protocols
Protocol

This is an optional step. However, we strongly recommend depositing your image data into a public repository, preferentially EBI's BioImage Archive (https://www.ebi.ac.uk/bioimage-archive/).

51. Zip each folder of collected raw data. Each folder is typically a single timepoint of a single plate with a size of 160 GB. Zipping can take up to 1 h per folder, and results in a zip file with a size of 60 GB. Running multiple processes at the same time can prolong the zipping time.
52. Register yourself with BioImage Archive: https://www.ebi.ac.uk/bioimage-archive/
53. Start a new submission in BioImage Archive. Fill all required fields.
54. Transfer all zip files, as well as the processed images (see optional: generation of processed images) into your personal area in BioImage Archive.
55. Create file lists for the raw data and the processed images.
56. Open the Anaconda Prompt, and execute:

```
> python create_filelist_raw_cq1.py METADATA_FILENAME
```

57. Check the Prompt for error messages and, if applicable, act on them.

   Note: Make sure to replace "METADATA_FILENAME" with the actual filename (see Section "document your data" above).

58. You will find a new folder called "filelists_raw" in the "root path", if all went well, and a file called "filelist_raw_data_cq1.xlsx" within that folder.
59. In the Prompt, execute:

```
> python create_filelist_processed_images_c1.py METADATA_FILENAME
```

60. Check the Prompt for error messages and, if applicable, act on them.

   Note: Make sure to replace "METADATA_FILENAME" with the actual filename (see "document your data" above).

61. You will find a new folder called "filelists_cq1" in the "root path", if all went well, and a file called "filelist_cq1_processed_images_all_experiments.xlsx" within that folder.
62. Create two Study Components in your submission to BioImage Archive.
63. Use the "filelist_raw_data_cq1.xlsx" to populate a Study Component about the raw data.
64. Use the "filelist_cq1_processed_images_all_experiments.xlsx" to populate a separate study component about the processed images.
65. Submit your study in BioImage Archive.

Evaluate your data

⏱ Timing: 1–2 h

This step describes the automated combination of data and metadata to derive the phenotypical characteristics and how to final evaluate each tested compound.

66. Open the Anaconda Prompt and execute:

```
> python merge_and_evaluate_cq1_csv_files.py METADATA_FILENAME
```

67. Make sure to replace "METADATA_FILENAME" with the corresponding filename (see "document your data").

68. Check the output for error messages and act accordingly.

69. You will find a new file called "cq1_evaluated_all.xlsx" in the "root path". Open it to manually inspect the results. An example of the layout can be found in the Table S3 (Zenodo: https://zenodo.org/record/7092795#.YyhponZByUk). The ratios used for validation were calculated by the following equations:

    1a = Ratio Hoechst High Intensity Objects
    = Cell_Count_Cell_Stats_HighIntObj/ Cell_Count_Cell_Stats
    1b = Ratio Normal Cells
    = Cell_Count_Cell_Stats_Normal / Cell_Count_Cell_Stats
    2a = Ratio Healthy Nuclei
    = Cell_Count_Cell_Stats_HealthyNuc/ Cell_Count_Cell_Stats_Normal
    2b = Ratio Fragmented Nuclei
    = Cell_Count_Cell_Stats_FragNuc/ Cell_Count_Cell_Stats_Normal
    2c = Ratio Pyknosed Nuclei
    = Cell_Count_Cell_Stats_PyknoNuc/ Cell_Count_Cell_Stats_Normal
    3a = Ratio Mitotic cells
    = Cell_Count_Cell_Stats_mitosis/Cell_Count_Cell_Stats_PyknoNuc
    3b = Ratio Apoptotic cells
    = Cell_Count_Cell_Stats_apoptosis/ Cell_Count_Cell_Stats_PyknoNuc
    4a = Ratio Intact membrane
    = Cell_Count_Cell_Stats_membrane intact/Cell_Count_Cell_Stats_HealthyNuc
    4b = Ratio Permeabilized membrane
    = Cell_Count_Cell_Stats_membrane permeab/ Cell_Count_Cell_Stats_HealthyNuc
    5a = Ratio Mitochondrial mass increased
    = Cell_Count_Cell_Stats_mito mass high/ Cell_Count_Cell_Stats_HealthyNuc
    5b = Mitochondrial mass normal
    = Cell_Count_Cell_Stats_mito mass normal/Cell_Count_Cell_Stats_HealthyNuc
    6a= Tubulin effect
    = Cell_Count_Cell_Stats_tubulin effect/Cell_Count_Cell_Stats_HealthyNuc
    6b = Tubulin normal
    = Cell_Count_Cell_Stats_tubulin normal/ Cell_Count_Cell_Stats_HealthyNuc

    a. Each row corresponds to the data of one compound at a certain timepoint
    b. All rows contain all information from the sheets in the metadata file, i.e., from the compound map, the compound batch, the compound, the imaging, and the experiment. This information can be used to filter down to entries of interest by standard spreadsheet software.
    c. The cell counts in each row are automatically compared to the mean of all wells in the same plate at the same timepoint marked as "blank" according to their "experimental type" (see "document your data"). This allows you to judge whether the phenotype is unique to the compound or generic to the experimental setup.
    d. As an example, we have implemented the ratio of the cell count per well for cells annotated as "healthy" and the mean of cell counts for cells annotated as "healthy" across all wells containing 0.1% DMSO (i.e., the "blank"). You can use this metric as a rough proxy of cytotoxicity.

70. Compare the values of interest for individual "blank" rows against the mean of all "blanks". This gives you a feeling for "natural distribution" of values for an "uninteresting" effect.

Note: Taking the sample data provided with this protocol, the range of values for the ratio of healthy cells (described in the step above) is approx. 0.6–1.2 for individual "blank" rows, with an average of approx. 0.9 and a standard deviation of 0.3. You can use this information to define thresholds to alarm you on "interesting" compounds.

STAR Protocols
Protocol

71. We also implemented a measure on "relative growth", comparing information from individual wells with compound-treated cells against the mean of "blank" wells.

   Note: More measures can easily be implemented in the Python script by interested researchers. You are also invited to contact the authors of this protocol.

   a. The "relative growth" was calculated by the following equation (Hafner et al., 2016):

$$relative\ growth = 2^{\frac{\log2\left(\frac{a_{tx}}{a_{t0}}\right)}{\log2\left(\frac{b_{tx}}{b_{t0}}\right)}} - 1$$

a = normal cell count compound.

b = normal cell count DMSO 0.1%.

tx = timepoint of interest.

t0 = timepoint after 0 h.

We defined threshold values (see Table 4), that will mark a compound with a comment. If a compound is marked in more than one cell line it should be flagged and the property should be evaluated in the context of the compound (target, concentration, IC50 etc.).

EXPECTED OUTCOMES

The aim of this protocol is the improved annotation of the cell health effect of a compound in a high-content manner. It is based on nuclear morphology and additional phenotypic characteristics such as tubulin effect, mitochondrial mass or membrane permeabilization. A detailed outcome exemplified by Itraconazol tested at 10 µM and 1 µM in comparison to DMSO 0.1% is shown in Figure 9.

After detection of cell bodies and nuclei using the CellPathfinder software (see data analysis CellPathfinder software) processed images (see optional: generation of processed images) or fluorescent images can be generated (Figure 9A). In order to assess the viability of the cells at a certain time point, the cell count per well (Figure 9B) or calculated growth rate (Figure 9C) can be shown. The growth rate serves as an assessment of whether a compound has cytostatic (GR 1-0) or cytotoxic effects (GR < 0) (Hafner et al., 2016). Further compound annotation is achieved by various additional gating steps (see data analysis CellPathfinder software). The ratios of compounds and controls can be represented in different ways e.g., pie charts (Figure 9D).

For automated compound annotation, certain thresholds regarding the gated properties were defined as shown in Table 4). It is nevertheless advisable to manually double check images and raw data before excluding a compound based on these thresholds. An example for compound annotation can be found in Figure 9E).

Table 4. Thresholds of compound properties

Property group	Property name	Threshold
Compound properties	Hoechst High Intensity Objects 1a	> 50%
Cell viability properties	Healthy Nuclei 2a	< 50%
	Fragmented Nuclei 2b	> 50%
	Pyknosed Nuclei 2c	> 50%
Phenotypical properties	Tubulin effect 6a	> 50%
	Mitochondrial mass 5a	> 50%
	Membrane permeabilized 4b	> 50%
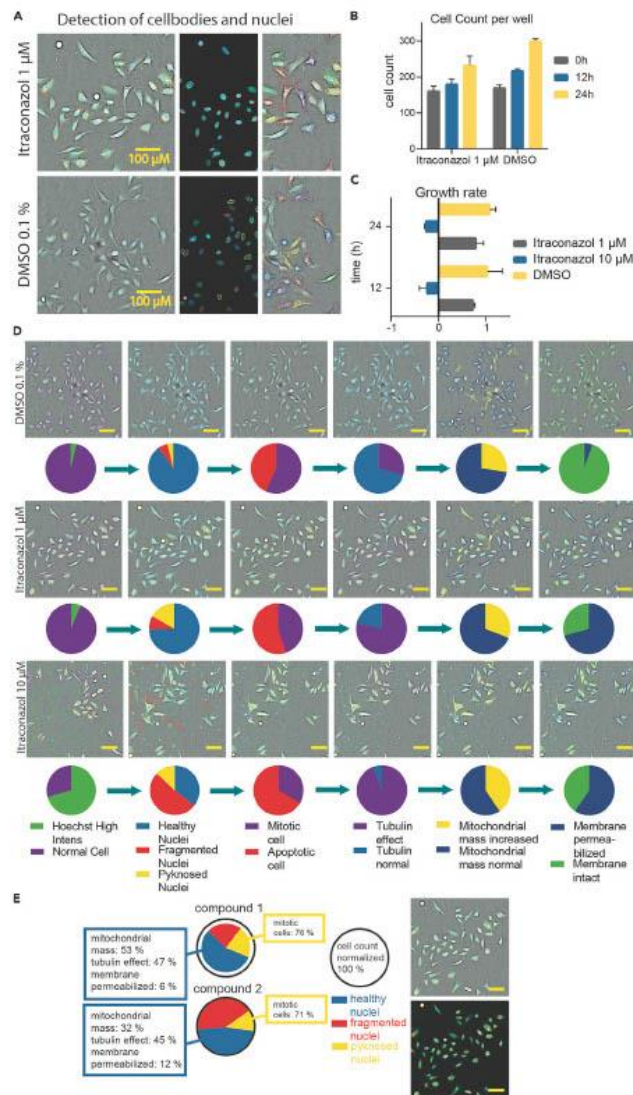
STAR Protocols
Protocol

Figure 9. Expected outcome after multiplex assay based on the example of Itraconazol

(A) Processed Image of stained (blue: DNA/nuclei, green: microtubule, red: mitochondria content, magenta: Annexin V apoptosis marker) U-2 OS cells after 24 h of exposure to Itraconazol (1 μM) and 0.1% DMSO and detection of nuclei and cell bodies after analysis using the CellPathfinder software.

(B) Detected cell count per well after 0 h, 12 h and 24 h of compound exposure (Itraconazol 1 μM) in comparison to cells treated with 0.1% DMSO. Error bars show SEM of two biological duplicates.

(C) Calculated relative growth called "growth rate" after 12 h and 24 h of compound exposure (Itraconazol 1 μM and DMSO 0.1%) to U-2 OS cells. Error bars show SEM of two biological duplicates.

(D) Gating of U-2 OS cells using the CellPathfinder analysis after exposure to 10 μM and 1 μM of Itraconazol in comparison to 0.1% DMSO after 24 h. Pie charts show the ratios of the different gating steps. Scale bars show 50 μm.

(E) Example figure for compound annotation after multiplex assay including gating ratios as well as two processed images of stained U-2 OS cells (blue: DNA/nuclei, green: microtubule, red: mitochondria content, magenta: Annexin V apoptosis marker).

Further examples of outcomes from this protocol are demonstrated in Tjaden et al.(Tjaden et al., 2022), where 215 compounds were tested successfully using the protocol described here.

LIMITATIONS

The described Multiplex assay contains some complex steps, which is why accurate performance of described working steps and data processing are necessary for data acquisition and analysis. Here, we describe the protocol for adherent U-2 OS cells. The protocol was also tested for the adherent cell lines HEK293T and MRC-9 fibroblasts (Tjaden et al.) (Tjaden et al., 2022). For other cell lines, adjustment of the measurement protocol as well as the analysis protocol for the CellPathfinder software might be necessary. The application of suspension cells or 3D culture was not tested and must therefore be carefully reviewed before using these cell lines.

Image acquisition of 384 well plates with multiple time points requires rather extended data storage capacity and computational processing power. We advise against saving the imaging data directly to a network connection, resulting in longer and not stable measuring time periods, which could bias the final results. Saving the data first locally on internal/external storage space followed by a data transfer to the suitable storage server is therefore suggested.

The measurement time for one plate determines the frequency of data acquisition possible. Therefore, time points cannot be chosen at too short intervals if a whole 384 well plate is measured. There are also temporal deviations between the first and the last measured well. We recommend therefore including the reference compounds at different plate areas to correct for this effect. For very high frequency data acquisition (<1–2 h) one can limit the number of imaged wells or adapt the image acquisition protocol.

In the workflow described here, all compounds are tested at 10 μM and 1 μM. Depending on target specificity and mechanism of action of the compounds, these concentrations can be adjusted.

The protocol is performed using a 10× magnification. For more detailed phenotypical readouts such as structure differentiation of tubulin or localization of mitochondria, higher magnification should be used. The magnification used in the protocol described here enables therefore only limited detection of intracellular spatial information.

The detection of cellular features in this protocol relies extensively on fluorescence intensity measurements. This is essential for the detection of nuclei using Hoechst33342 which enables the further gating steps based the shape of cellular nuclei. Compounds that interfere with light absorption such as fluorescent compounds, may interfere with this protocol and result in a high ratio of "Hoechst High Intensity Objects".

Compounds, insoluble in DMSO and thus precipitated in their DMSO stocks can result in false negative outcomes (see problem 4). Chemical quality control of the DMSO stocks of tested compound using LC-MS can help detect and account for this occurrence.

TROUBLESHOOTING

Problem 1

A different cell line than U-2 OS is used and not properly detected and/or gated.

If different cell lines are used, the analysis protocol for U-2 OS cells is no longer applicable due to different cell morphology.

Potential solution

The measurement protocol as well as the analysis protocol should be adjusted accordingly. We tested two other adherent cell lines (HEK293T and MRC-9 fibroblasts), where we could see that mostly the cell body detection and the machine-learning based algorithm must be changed due to the different overall cellular morphology of the different cell lines. For the adherent cell lines HEK293T and MRC-9 fibroblasts, the adjusted protocol can be found here: Zenodo: https://doi. org/10.5281/zenodo.6415330. Other cell lines, especially suspension cells, have not been optimized or tested (see limitations).

Problem 2

Cell seeding density is not optimal. Either too little cells are observed in each image or high confluency impedes the segmentation of individual cells.

Potential solution

Optimize the cell seeding density as described in steps 2–5 in section optimize cell seeding density. Proper assessment of the cell viability and cell passaging at a confluence not higher than 90% can increase robustness and reproducibility of this assay. As mentioned above an optimal confluence 24 h after seeding should be 90%. If the cell density cannot be optimized by the cell seeding number, the imaging time points could be prolonged or shortened to obtain better results.

Problem 3

Evaporation of the medium.

Certain wells, especially on the plate edges may contain less liquid and a higher apoptosis rate can be observed in control wells located at the edge of the 384 well plate.

Potential solution

Increased evaporation of cell medium can affect the outcome of the experiment by changing the composition of medium components and increasing the concentration of tested compounds.

As shown in the layout (see Figure 2), it is best practice to exclude the outer wells of the 384 well plate from the measurement and fill those with 100 μL PBS buffer or water (see addition of fluorescent dyes and cell seeding).

Problem 4

Compounds are not soluble in DMSO and/or precipitation of compounds on LDV source plate.

Compounds that are insoluble in DMSO or less soluble than others can either not be dispensed by the used ECHO Labcyte protocol to the cells or can precipitate in the source plate.

Potential solution

DMSO solubility of compounds should be tested beforehand. If not soluble in DMSO, try diluting in water and pipet by hand or with ECHO Labcyte protocol for aqueous solvents (384LDV_aq). Look by eye to see if the compound precipitated on the Echo plate before pipetting. LC-MS

STAR Protocols
Protocol

can help detect and account for this occurrence. Frequent freeze-thaw-cycles can increase this problem.

Problem 5
Cells are not evenly distributed within the well.

An uneven distribution of cells can lead to suboptimal cell growth under the assay conditions or a falsely overestimation of the compound toxicity. Additionally, a lower number of cells compromises the statistical analysis and can impede the clarity of the results.

Potential solution
Before testing the cells, make sure that the cells have the optimal cell seeding density and are detached by trypsin treatment to avoid clusters (see optimize cell seeding density). Gently mix the cells before seeding to ensure an even distribution of the cells in the culture medium. The waiting step described in 6 is important to avoid any strong convection currents within the wells due to the different temperature of the medium and the surface of the plate. The waiting period allows cells to start attaching themselves to the plate surface and makes them less susceptible to this phenomenon when placed subsequently in the incubator.

If the majority of the cells are found on the outer rim of the wells of an already prepared plate, one can change the image acquisition spot of the well at the CQ1 protocol to a spot less centered.

Problem 6
The focus of CQ1 protocol is set incorrectly.

If the focal plane of the microscope has not been properly adjusted, cells may not be detected and unwanted errors due to focusing on dirt on the plate may limit the measurement. This can happen when a different plate, not implemented in the CQ1 software is used, or if the plate is uneven or if there is evaporation on the plate.

Potential solution
The plate dimensions of the used plates should be implemented in the CQ1 software. Make sure you can adopt the temperature of the microscope to avoid condensation occluding the imaging and to prevent focus problems due to evaporation on the plate lid. To set the right focus, test the autofocus of the CQ1 on your plate using Channel 1. If the focus area differs from the already set focus, change the protocol accordingly.

Problem 7
The results of the reference compounds after the pre-analyzing using the machine learning-based algorithm differed by more than 10%.

After the pre-analysis the results from the reference compounds should not differ by more than 10%. This can happen when the normal growth conditions are affected e.g., due to evaporation (see problem 3), not evenly distribution (see problem 5), different cell morphology due to overgrowing of the cells or the fluorescence intensity levels differed from the training set. If so, the machine-learning-based algorithm should be adjusted accordingly.

Potential solution
The machine learning-based algorithm has to be trained for every new experiment. The same algorithm can then be used for the following experiments with the same conditions. To check if the analysis worked appropriately, the reference compounds are added in quatruplicates to double-check your results (see steps 36–39 in data analysis CellPathfinder software). If the results of the

reference compounds differ from each other more than 10%, the machine learning algorithm should be trained again with new example images. For this, follow steps 34–36 of Data analysis CellPathfinder Software. Example training images can be found in Figure 7.

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Susanne Müller-Knapp (susanne.mueller-knapp@bmls.de).

Materials availability
This study did not generate new unique reagents.

Data and code availability
The images and raw datasets generated during this study are available at BioImage Archive: https://www.ebi.ac.uk/biostudies/studies/S-BIAD145. Analysis data and codes reported in this paper can be found at Zenodo: https://doi.org/10.5281/zenodo.6325622. All supplemental data have been deposited to Zenodo database under the following link: Zenodo: https://zenodo.org/record/7092795#.YyhponZByUk.

Original data for figures in the paper are available at Tjaden et al. (Tjaden et al., 2022).

AUTHOR CONTRIBUTIONS

Conceptualization, A.T., S.M., and M.S.; Methodology, A.T. and M.S.; Investigation, Validation and Formal Analysis, A.T.; Data Curation, R.T.G.; Resources, S.M. and S.K.; Writing – Original Draft, A.T. and R.T.G.; Writing – Review & Editing, M.S., S.M., and S.K.; Funding Acquisition, S.K.; Supervision, M.S., S.M., and S.K. All authors have read and agreed to the published version of the protocol.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR Protocols
Protocol

REFERENCES

Al-Aamri, H.M., Ku, H., Irving, H.R., Tucci, J., Meehan-Andrews, T., and Bradley, C. (2019). Time dependent response of daunorubicin on cytotoxicity, cell cycle and DNA repair in acute lymphoblastic leukaemia. BMC Cancer 19, 179.

Brayton, C.F. (1986). Dimethyl sulfoxide (DMSO): a review. Cornell Vet. 76, 61–90.

Bruno, S., Ardelt, B., Skierski, J.S., Traganos, F., and Darzynkiewicz, Z. (1992). Different effects of staurosporine, an inhibitor of protein kinases, on the cell cycle and chromatin structure of normal and leukemic lymphocytes. Cancer Res. 52, 470–473.

Fokas, E., Prevo, R., Pollard, J.R., Reaper, P.M., Charlton, P.A., Cornelissen, B., Vallis, K.A.,

Hammond, E.M., Olcina, M.M., Gillies McKenna, W., et al. (2012). Targeting ATR in vivo using the novel inhibitor VE-822 results in selective sensitization of pancreatic tumors to radiation. Cell Death Dis. 3, e441.

Hafner, M., Niepel, M., Chung, M., and Sorger, P.K. (2016). Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. Nat. Methods 13, 521–527.

Musa, M., Kannan, T.P., and Mustafa, S. (2013). Cell proliferation study of human osteosarcoma cell line (U2OS) using Alamar Blue assay and live cell imaging. IOSR J. Dental Med. Sci. 8, 60–65.

Sánchez-Martínez, C., Gelbert, L.M., Lallena, M.J., and de Dios, A. (2015). Cyclin dependent kinase

(CDK) inhibitors as anticancer drugs. Bioorg. Med. Chem. Lett. 25, 3420–3435.

Styrt, B., Johnson, P.C., and Klempner, M.S. (1985). Differential lysis of plasma membranes and granules of human neutrophils by digitonin. Tissue Cell 17, 793–800.

Tjaden, A., Chaikuad, A., Kowarz, E., Marschalek, R., Knapp, S., Schröder, M., and Müller, S. (2022). Image-Based annotation of chemogenomic libraries for phenotypic screening. Molecules 27, 1439.

Wang, T.H., Wang, H.S., and Soong, Y.K. (2000). Paclitaxel-induced cell death: where the cell cycle and apoptosis come together. Cancer 88, 2619–2628.

Appendix C: Characterization of Cellular Viability Using Label-Free Brightfield Live-Cell Imaging

Contribution Menge, A.: Conceptualization, Original Draft, Writing – Review & Editing

Chapter 6

Characterization of Cellular Viability Using Label-Free Brightfield Live-Cell Imaging

Lewis Elson, Amelie Tjaden, Stefan Knapp, and Susanne Müller

Abstract

In recent years, the assembly and annotation of chemogenomic libraries have gained interest by the phenotypic screening community. Apart from basic annotations of the compound potency and selectivity, these compound libraries benefit in particular from annotation regarding the effect of the inhibitors on cellular viability to distinguish between on-target effects of a compound and unspecific cytotoxicity. Here, we provide a protocol to determine viability as a first determinant in compound quality control, using the Incucyte live-cell imaging system. The compounds are classified according to their calculated growth rate to determine a cytotoxic, cytostatic, or healthy outcome. All compounds affecting the growth rate can be further evaluated regarding their specific effects on cell health in a high-content live-cell multiplex assay, described in Chapter 5.

Key words Viability assay, Phenotypic screening, Cytotoxicity, Growth rate

1 Introduction

The use of well-characterized screening libraries has garnered a rise in popularity over recent years. These libraries ideally consist of compounds with high target specificity, so-called chemical probes, or compounds with a narrower (usually restricted to a few members of a family) but not exclusive selectivity profile, so-called chemogenomic compounds [1]. The latter has seen increased interest due to the current lack of readily available chemical probes [2]. With the resurgence of phenotypic screening over traditional target-based drug discovery efforts, as it offers the advantage of not requiring the full understanding of a specific mode of action and being more physiologically relevant [3], these libraries can offset the lack of mechanistic insight that can complicate hit validation, often associated with such approaches [1]. While these libraries offer a way to

validate targets associated with a specific phenotype, it is important to note that non-specific compound toxicity will impact phenotypic readouts and thus the ability to interpret the data and associated phenotypes with the relevant molecular targets [1]. With this in mind, we propose that each compound should be characterized based on its effect on cells, in particular its impact on cell viability.

A frequently employed way to assess and profile small molecules based on their effects on cellular viability are colorimetric or fluorescent-based assay systems such as alamarBlue™ or the tetrazolium dye 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) [4]. In contrast to these plate reader-based approaches, live-cell imaging systems can assess cell viability without the use of additional reagents that might in turn affect cell growth. Here, cellular confluency is measured over a defined period of time before and after compound treatment. For adherent cells, confluency refers to the percentage of the culture vessel's surface area that is covered by a layer of cells. Suspension cell confluency can be determined in a similar manner (how crowded a defined area is), but is usually assessed via medium turbidity and colorimetric assessment.

The confluency parameter is a simple and cost-effective way to quantify a compound's effect on the cell. To validate the confluency independent of both cell division rates and cell seeding densities while also including time-dependent changes of compound effects, test compounds can be evaluated based on their growth rate. The growth rate is a normalized parameter (against the negative control) that provides a quantitative measure by incorporating initial cell confluency allowing for the characterization of either cytotoxic or cytostatic effects [5].

Calculated growth rate values lie in a range between "−1" and "1." Between "0" and "1," a compound shows partial inhibition of the cell growth, "0" defines a cytostatic effect, while values between "0" and "−1" define a cytotoxic effect (Fig. 1a). A value of "1" matches the value for the negative control and thus denotes no effect. Compounds that are classified as cytotoxic have growth rates lower than 1 (Fig. 1b) or appear to show a distinct morphological effect (Fig. 2) are considered as "hits" and can be further evaluated in other assays, if appropriate, such as a high-content live-cell multiplex assay (Chapter 5) which aims to elucidate the particular effect of the compound within the cell.

Several live-cell imaging systems are available, such as the CELLCYTE X™ Live-Cell Imager (CYTENA), which can be used for the assay; however, here, we describe the workflow for a one-shot assay capable of characterizing general cellular viability using the Incucyte Live Cell imaging system. With this protocol, screening of 308 small molecules (including controls) on a 384-well microplate format is possible. The main procedures
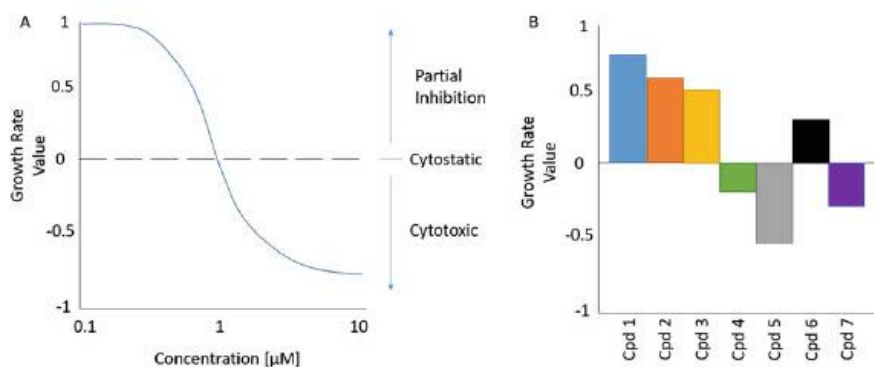
Fig. 1 Graphical illustration of growth rate. (**a**) Example of a dose–response curve for a compound and potential outcomes at differing concentrations. As the concentration of a compound increases it is more likely to exhibit cytotoxic effects and presents graphically as an inverse sigmoid function. (**b**) Comparison of compound effect at a fixed concentration. With this format, it is possible to differentiate between the effect type of each compound and base a selection from the information presented
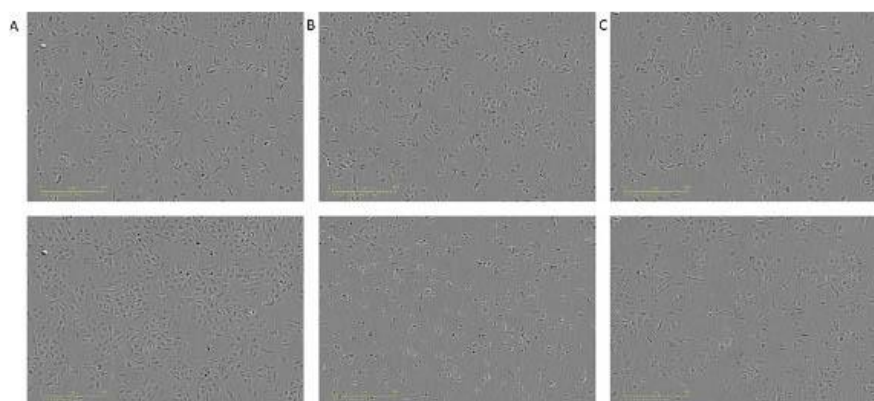
Fig. 2 Brightfield images of U2OS cells incubated with DMSO (**a**) Daunorubicin (**b**) and Milciclib (**c**) at 0h (top) and 24h (bottom) showing healthy, cytotoxic, and cytostatic effects, respectively

involve cell seeding, Incucyte image acquisition setup, image analysis and export and growth rate calculation for compound classification and viability characterization. The approach described here is transferable to other systems, but will require adaptations, depending on the instrument used.

It is recommended to run biological replicates for this assay due to the lack of technical duplicates.

78   Lewis Elson et al.

Example images from previous experiments are available for reference on the BioImage Archive database (https://www.ebi.ac.uk/biostudies/studies/S-BIAD145#) [9].

2   Materials

All solutions must be prepared under sterile conditions and kept so; all work involving live samples requires sterile working. Use cell lines that have not undergone a substantial number of passages to ensure genomic integrity [6]. Cell culture reagents including medium, washing solution, and reagents for cells passaging should be stored at 4 °C and be warmed up to 37 °C prior to use. Compounds should be prepared as aliquots and can be stored sealed at room temperature (RT) for approximately 3–6 months. Try to avoid freeze–thaw cycles [7] and ensure that aliquots are soluble at the stored concentration and are not precipitated prior to use. For longer term storage, stock solutions should be stored at − 20 °C.

2.1   Cell Lines

This protocol describes a workflow using three adherent cell lines. Suspension cell lines will require a different workflow. The cell lines described here are well-characterized and represent transformed (HEK293T, U2OS) as well as non-transformed cell lines (MRC-9). All cell lines are categorized as Biosafety Level 1 and are kept incubated at 37 °C with 5% CO_2.

1. HEK293T (DMSZ: ACC635)—human embryonal kidney cells, adherent and epithelial in morphology, grown as monolayer, DMEM.
2. U-2 OS (ATCC: HTB-96)—human osteosarcoma, adherent with epithelial morphology, grown as monolayer, DMEM.
3. MRC-9 (ATCC: CCL-212)—human lung fibroblasts, adherent with fibroblast morphology, grown as monolayer, EMEM.

2.2   Reagents and Perishables

As mentioned previously in Subheading 2.1, the medium listed here is specific to the cell lines we present as an example. Please ensure you use correct media and supplementation for your cell lines of choice.

1. Dulbecco's modified Eagle's medium (DMEM) (Thermo Fisher: 11965084): supplement with 10% fetal bovine serum (Thermo Fisher: 26140079) and 1% penicillin–streptomycin (Thermo Fisher: 15140122).
2. Eagle's minimum essential medium (EMEM) (ATCC:30-2003): supplement with 10% fetal bovine serum (Thermo Fisher: 26140079) and 1% penicillin–streptomycin (Thermo Fisher: 15140122).

3. Cell washing solution: Dulbecco's phosphate-buffered saline 1× (DPBS): without calcium or magnesium (Thermo Fisher: 14190144).

4. Dissociation reagent: trypsin-EDTA 1× (0.05%) (Thermo Fisher: 15400054).

5. 0.4% trypan blue solution (Thermo Fisher: 15250061).

6. Cell culture flask: T75 is sufficient (75 cm^2/250 mL) (Greiner: 658170).

7. Serological pipettes (Eppendorf: 0030127722) and PIPET-BOY (Integra 155017).

8. Mechanical pipettes (Eppendorf: 3123000918).

9. Greiner Bio-One 384-well standard CELLSTAR polystyrene microplate (Greiner: 781091) or compatible 384 microplate.

10. 15-mL and 50-mL tubes (Greiner: 188261/2120261).

11. Disposable reservoirs (Thermo Fisher: 95128095).

2.3 Reference Compounds

Reference compounds should cover outcomes that include cytotoxicity, cytostaticity, cell lysis/permeabilization, and having no observed effect on cell growth at the concentrations used. Described below are reference compounds commonly used as controls for the experiment. These compounds are tested at 10 μM (0.1% v/v for DMSO) and are distributed evenly throughout the plate. If compounds are dissolved in water, exchange DMSO for PBS/Water.

1. Daunorubicin—positive [10].

2. Staurosporine—positive [11].

3. Digitonin—positive [12].

4. Milciclib—positive [13].

5. Bromosporine—positive [14].

6. JQ1—positive [15].

7. Paclitaxel—positive [16].

8. Torin—positive [17].

9. DMSO—no effect [18].

2.4 Instruments and Related Software

1. Incucyte® SX1, S3 or SX5 Live-Cell Analysis System (Sartorius AG), or similar device.

2. Incucyte® Base Analysis Software.

3. Centrifuge.

4. CO_2 incubator.

5. Light microscope.

6. Cell counting device: automated cell counter (i.e., TC20 Bio-Rad) or hemocytometer (Fisher Scientific).

3 Methods

Carry out all procedures at room temperature and under sterile conditions using aseptic technique unless specified otherwise. Seeding density for different cell lines will need to be established prior to the experiment (*see* **Note 1**).

3.1 Cell Seeding

1. Aspirate medium from cell culture flask by angling the flask and add 5 mL 1× DPBS to the corner of the flask; rinse cells with the solution.

2. Aspirate DPBS from the flask and add 3 mL of 1× trypsin-EDTA; return flask to the incubator for 1–2 min (dependent on cell line).

3. Remove flask from incubator and determine the detachment of cells via light microscopy.

4. Once the cells are detached, add 7 mL of culture medium to neutralize trypsin activity and transfer dilution to a 15-mL tube.

5. Spin down at 200× g for 5 min.

6. Aspirate supernatant without disturbing the pellet and resuspend in fresh medium.

7. Create a 1:1 dilution of cell suspension and 0.4% trypan blue (10 μL of cells in 10 μL trypan blue), and add 10 μL to the counting slide and insert into the automated cell counter (or count using a hemocytometer).

8. Note the cell number and viability (viability above 95% is acceptable). Create a cell suspension for a desired cell number per well in a chosen volume resulting in a confluency value of approximately 40%. Cell seeding density (cell/mL) is calculated via the formula as follows:

$$(1000/\text{volume per well}) \times \text{cells per well}.$$

For example, seeding 1500 cells per well in 50 μL: $(1000/50) \times 1500 = 30{,}000$ cells/mL solution.

9. Cell numbers for cell lines described above consist of the following:
   (a) HEK293T—1750 cells/well.
   (b) U2OS—1500 cells/well.
   (c) MRC-9—1250 cells/well.

10. Add × µL of your cell solution (volume calculated from **step 8**) to each well of the plate (avoiding the outer perimeter, *see* **Note 4**), allow to rest for up to 20 min at room temperature (*see* **Note 2**).

11. Remove any air bubbles present in the wells as they can interfere with the sedimentation and visualization of the cells (*see* **Note 3**).

12. Place in incubator at 37 °C 5% CO_2 overnight.

13. The following day check your plates and observe the status of the cells. They should be dispersed evenly throughout each well and be ~40% confluent; adherent cells should be adhered to the plate. From here you can set up the Incucyte image acquisition.

3.2 Incucyte Image Acquisition Setup

The following setup is specifically tailored to the Incucyte live cell imaging system; other devices will require adaptation to the protocol.

The Incucyte software is divided into two main parts: acquiring scans and viewing and analysis of the scans. Acquisition allows for vessel configuration, scheduling, image acquisition, and the storage of vessel data into a database. Viewing and analysis involve measuring and assessing the acquired vessel data. Any general user queries regarding the software can be answered by the user manual [8].

1. Open the Incucyte imaging software and enter your credentials.

2. Select **Schedule** to enter the acquisition window.

3. Select **Launch Vessel Wizard**, the **Add Vessel** wizard opens.

4. Select **Scan on Schedule** and click **Next**, the **Create Vessel Page** opens.

5. The **Create or Restore Vessel Page** provides four options as follows:

   (a) **New**—creates an entirely new vessel to scan.

   (b) **Copy Current**—creates a new vessel by copying a vessel from the current schedule.

   (c) **Copy Previous**—creates a new vessel by copying a previously scanned vessel.

   (d) **Add Scans (Restore)**—restores a previously scanned vessel for additional scanning.

6. Under **Create Vessel** select **New**, this will open the **Scan Type** page.

7. On the **Scan Type** page select **Standard**, this will open the **Scan Settings** page.

8. Specify the image channel **Phase** and objective to **10×**. Click **Next** and the **Vessel Type Search** page will open.

9. On the **Vessel Type** window, you will need to select the correct plate; for our assay, we use Greiner 384 Serial Number 781091 (other plate types can be used but ensure they are compatible and the correct plate is selected). Select **Next** and the **Vessel Locator** page opens.

10. **Select a location** in which the plate will be placed, click **Next**, and the Scan Pattern page will open.

11. The **Scan Pattern** page displays a vessel map which will show available wells to be scanned based on the plate you have selected in the **Vessel Type** window. Drag and drop to select the inner 308 wells and specify the number of images per well as 1 (384 plates only allow 1 image per well) (*see* **Note 4**). Once selected, click **Next** to bring up the **Vessel Information** window.

12. **Vessel Information** requires a **Name for the plate**, and enter cell type and passage at your leisure. A plate map can be created manually or imported to define each well with compound and cellular conditions. Click **Next** and the **Analysis Setup** window will appear.

13. Click **Next** on the **Analysis Setup** window, and analysis will be defined later.

14. Set the scanning schedule by dragging the bar across the timeline and set the scan time to 6-h intervals. In the **Stop Scan** section, select after 24 h (time point measurements can be changed to suit the user's preference). Click Next.

15. The **Summary Page** will outline the experimental run, and click **Add to schedule**.

3.3 Incucyte Plate Run

1. Insert the plate into the cassette that was designated earlier (*see* **step 10** in subheading 3.2) to take the blank measurement (*see* **Notes 5, 6, 10** and **11**). The blank is required for calculation of the growth rate (*see* subheading 3.6).

2. Manipulate the schedule via the timeline on the acquisition window to scan the plate within the next 5–10 min. You can see the progress of the scan at the top right of the acquisition window.

3. Once scanned, select the **View** tab and double-click on the scan; check that each well is visible and there are no resolution problems, smears, or scratches (*see* **Notes** 7 and **8**).

4. Remove the plate from the device, add test and reference compounds to the plate to a final concentration of 10 μM. (10 μM is a good benchmark to expect toxicity if it were to occur). Other concentrations may be used depending on the expected potency of your compounds.

5. Return the plate to the Incucyte, and manipulate the schedule via the timeline on the acquisition window to scan the plate within the next 5 min (*see* **Note 9**).

6. Once scanned, select the **View** tab and double-click on your scan, and ensure that each well is visible and there are no resolution problems, smears, scratches, or apparent precipitation of compounds (*see* **Notes 5** and **6**).

7. Allow the plate to run for full 24 h duration.

3.4 Incucyte Plate Analysis

1. Select your plate via the **View** tab.

2. A plate overview will appear with options on the left side of the window. Check each time point to ensure all wells are in focus.

3. Select **Launch Analysis** from the analysis toolbar.

4. Select **Create a New Analysis Definition**.

5. Select **Basic Analyzer**.

6. Select the **Image Channels**, and this assay only uses **Phase**.

7. Select images for the software to train with, and we suggest to include wells that consist of different morphologies while incorporating a negative and positive control wells for good coverage.

8. Define the analysis parameters; this process will involve trial and error to create a fitting mask (*see* **Notes 12** and **13**).
   For HEK293T/U2OS/MRC-9, the following is recommended as follows:

   (a) Hole fill (μm): 100/100/200.
   (b) Adjust size (pixels): $-1/-1/-1$.
   (c) Area (μm^2): Select Min and insert 100/100/200.
   (d) Eccentricity: Select Min and insert 0.3–0.5/select Min and insert 0.3–0.5/N/A.

9. Select the **Preview All** function to visualize the defined mask for all selected wells. A well-fitting mask outlines cell morphology well while excluding debris, if acceptable click **Next**.

10. On the **Scan Times** and **Wells** window select **all time points and wells**, click **Next**.

11. To Save and Apply the analysis definition input a **Definition Name**, analysis notes are not required. Click **Next**.

12. Review and verify the summary information, and click **Next** to apply the analysis definition. The software will now run the analysis; you can view the progress of this via the status tab.

3.5 Data Export and Analysis

1. After the vessel data has been analyzed, the vessel window will now have the associated analysis masks that were applied.

2. Open the vessel and click the **Graph Metrics** icon to open the Graph Metrics window.

3. Select the **Confluence (%)** metric from the metrics pane.

4. Select all time points and wells you measured. From here you can select as follows:

    (a) **Microplate Graph**—Shows an overview of every well analyzed and their confluency against time point.

    (b) **Graph**—each well is plotted on a standard xy plot.

    (c) **Export**—export the values to a third-party software.

5. Select **Export** and specify the format in which the data should be exported via the layout and destination can be specified through three options as follows:

    (a) **Clipboard**—to be able to manually paste into a third-party software program such as excel.

    (b) **All scans in one file**—specify where the scans should be stored as a combined file.

    (c) **Each scan in a separate file**—specify where each scan should be stored as individual files.

    (d) Select other options under **Other Options**.

3.6 Growth Rate Calculation

Determining the growth rate is calculated according to the formula [5]:

$$GR(c) = 2^{\frac{log2(x(c)/x_0)}{log2(x_{ctrl}/x_0)}} - 1$$

$x(c)$ is the treated cell count.

x_{ctrl} is the control cell count.

x_0 is the cell count prior to treatment (Blank).

The workflow and an example of how to calculate this value are illustrated below:

Blank value—12.

After 24 h:

Test compound—13.

Negative control (each well)—25, 26, 27, 28, 29.

To calculate the growth rate:

1. Calculate the negative control average confluency value for the chosen time point.

$$(25 + 26 + 27 + 28 + 29)/5 = 27.$$

2. Take the test compound, blank and negative control confluency values, and input them into the calculation below:

$$GR(c) = 2^{\frac{log2(x(c)/x_0)}{log2(x_{ctrl}/x_0)}} - 1$$

$x(c)$ is the treated cell count—this is the compound confluency value.

x_{ctrl} is the control cell count—this is the average negative control confluency value.

x_0 is the cell count prior to treatment—this is the blank confluency value.

3. This should be the base formula with the relevant values as follows:

$$GR(c) = 2^{\frac{log2(13/12)}{log2(27/12)}} - 1$$

13 is the confluency value for the test compound.

12 is the confluency value for the blank measurement.

27 is the confluency value for the negative control.

4. Calculate the first set of brackets. Note - values have been rounded to three significant figures for simplicity.

$$GR(c) = 2^{\frac{log2(1.08)}{log2(2.25)}} - 1$$

5. Find the log base 2 (log2) of the result.

$$GR(c) = 2^{\frac{log2(1.08)}{log2(2.25)}} - 1$$

$$GR(c) = 2^{\frac{0.111}{1.17}} - 1$$

6. Calculate the remaining values.

$$GR(c) = 2^{0.0949} - 1.$$

7. Finish the calculation.

$$GR(c) = 0.0680.$$

To minimize the risk of error, it is recommended to use third-party software for calculation of the growth rate, illustrated below is an excel formula that can be used for calculation:

$$= (2\ ((\text{LOG}(((x(c)/x_o),2)/\text{LOG}((x_{ctrl}/x_o),2))))-1.$$

$x(c)$ is the treated cell count—this is the compound value.

x_{ctrl} is the control cell count—this is the average negative control value.

x_0 is the cell count prior to treatment—this is the blank value.

4   Notes

1. Different cell lines will require individual optimization to determine a good seeding confluency; for example, we find that most adherent cells work well between 1500 and 2000 cells per well in a 384-well plate. It is important to have sufficient cells to influence growth but not too many so that the initial confluency value is too high.

2. Allowing the plate to sit at room temperature for 20min minimizes the impact of convection currents when moved to the incubator, keeping the cells from aggregating at the edge of the wells.

3. When pipetting your plates, avoid creating bubbles as much as possible, as they can interfere with the attachment and distribution of your cells. If bubbles are formed, it is possible to use a so-called debubbler to remove them. This resembles a laboratory wash, a nozzle inserted into an empty plastic bottle which you are able to squeeze to expel liquid for cleaning purposes (as an example Thermo: 2401-0125). When the bottle is empty and squeezed, air is expelled and will remove any bubbles that have formed within the wells.

4. To avoid the so-called edge-effect, do not use the outer perimeter of the plate, and it can be used as protection by being filled with DPBS or media.

5. When inserting your plate into the Incucyte cassette, do so using a slide motion. Do not press the plate down into the cassette as this can damage the motor of the drawer.

6. Use fresh plates or those that have been sealed previously. Plates that have been standing around can accumulate dust and other debris that will interfere with the analysis software.

7. Avoid touching or abrasively wiping the bottom of the microplate as much as feasibly possible. It can result in smearing and scratches which are detected by the Incucyte optics and analysis software, leading to unreliable confluency values.

8. The Incucyte may have a focusing problem from time to time. If your scan appears blurry, simply rescan at a later time, it should rectify itself.

9. While the Incucyte will display a warning when scheduling additional experiments too close to one another, it is advised to leave at least 5 min before each scan regardless. This ensures that different scans will not overlap and cause a loss of data.

10. The Incucyte will generate heat while working; this is unavoidable. This temperature increase can have an impact on your assay. It is often recommended to set the temperature of the incubator your Incucyte is housed in to 0.5°C lower than 37°C to offset this change.

11. Condensation can build up on the plate's lid when placed into the Incucyte. It is recommended to allow the plate to acclimatize for up to 10min before beginning a scan.

12. Analysis parameters will be cell line specific; different morphologies will dictate the criteria you set.

13. While the parameters here will minimize most unwanted debris from being detected, be aware some compounds can cause a significant increase in confluency values despite causing cell death. An example would be digitonin, a non-ionic detergent that permeabilizes the membrane and causes leakage of cytosolic constituents resulting in a haze effect. It is recommended to assess each image to identify whether the confluency value can be trusted.

Acknowledgement

88   Lewis Elson et al.

References

1. Tjaden A, Chaikuad A, Kowarz E, Marschalek R, Knapp S, Schröder M, Müller S (2022) Image based annotation of chemogenomic libraries for phenotypic screening. Molecules 27:1439

2. Müller S et al (2022) Target 2035 – update on the quest for a probe for every protein. RSC Med Chem 13:13–21

3. Moffat JG, Vincent F, Lee JA, Eder J, Prunotto M (2017) Opportunities and challenges in phenotypic drug discovery: an industry perspective. Nat Rev Drug Discov 16:531–543

4. Wagner BK, Schreiber SL (2016) The power of sophisticated phenotypic screening and modern mechanism-of-action methods. Cell Chem Biol 23:3–9

5. Hafner J, Niepel M, Chung M, Sorger PK (2016) Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. Nat Methods 13:521–527

6. Hughes P, Marshall D, Reid Y, Parkes H, Gelber C (2018) The costs of using unauthenticated, over-passaged cell lines: how much more data do we need? BioTechniques 43. https://www.future-science.com/doi/10.2144/000112598

7. Kozikowski BA et al (2003) The effect of room-temperature storage on the stability of compounds in DMSO. J Biomol Screen 8:205–209

8. Sartorius (2022) Incucyte live-cell analysis systems user manual. https://www.sartorius.com/download/1087348/incucyte-live-cell-analysis-systems-user-manual-en-l-8000-04-1%2D%2Ddata.pdf. Accessed 17 June 2022

9. EMBL-EBI (2022) BioImage Archive https://www.ebi.ac.uk/bioimage-archive/. Accessed 4 July 2022

10. Al-Aamri H et al (2019) Time dependent response of daunorubicin on cytotoxicity, cell cycle and DNA repair in acute lymphoblastic leukaemia. BMC Cancer 19. https://doi.org/10.1186/s12885-019-5377-y

11. Bruno S et al (1992) Different effects of staurosporine, an inhibitor of protein kinases, on the cell cycle and chromatin structure of normal and leukemic lymphocytes. Cancer Res 51:470–473

12. Seina J et al (2001) Cytotoxicity of digitoxin and related cardiac glycosides in human tumor cells. Anti-Cancer Drugs 12:475–483

13. Sanchez-Martinez C, Gelbert LM, Lallena MJ, Dios A (2015) Cyclin dependent kinase (CDK) inhibitors as anticancer drugs. Bioorg Med Chem Lett 25. https://doi.org/10.1016/j.bmcl.2015.05.100

14. Picaud S et al (2016) Promiscuous targeting of bromodomains by bromosporine identifies BET proteins as master regulators of primary transcription response in leukemia. Sci Adv 2. https://doi.org/10.1126/sciadv.1600760

15. Wen N et al (2019) Bromodomain inhibitor JQ1 induces cell cycle arrest and apoptosis of glioma stem cells through VEGF/PI3K/AKT signalling pathway. Int J Oncol 55. https://doi.org/10.3892/ijo.2019.4863

16. Wang TH et al (2000) Paclitaxel-induced cell death. Am Cancer Soc J 88. https://doi.org/10.1002/1097-0142(20000601)88:11<2619::AID-CNCR26>3.0.CO;2-J

17. Francipane MG, Lagasse E (2013) Selective targeting of human colon cancer stem-like cells by the mTOR inhibitor Torin-1. Oncotarget 4. https://doi.org/10.18632/oncotarget.1310

18. Violante GA et al (2002) Evaluation of the cytotoxicity effect of dimethyl sulfoxide (DMSO) on CaCo$_2$/TC/colon tumor cell cultures. Biol Pharmaceut Bull 25:1600–1603

Appendix D: Annotation of the Effect of Chemogenomic Compounds on Cell Health Using High-Content Microscopy in Live-Cell Mode

Contribution Menge, A.: Conceptualization, Methodology, Investigation, Validation and Formal Analysis, writing – Original Draft, writing, review and editing.

Chapter 5

Annotation of the Effect of Chemogenomic Compounds on Cell Health Using High-Content Microscopy in Live-Cell Mode

Amelie Tjaden, Stefan Knapp, and Susanne Müller

Abstract

The characterization of chemogenomic libraries with respect to their general effect on cellular health represents essential data for the annotation of phenotypic responses. Here, we describe a multidimensional high-content live cell assay that allows to examine cell viability in different cell lines, based on their nuclear morphology as well as modulation of small molecules of tubulin structure, mitochondrial health, and membrane integrity. The protocol monitors cells during a time course of 48 h using osteosarcoma cells, human embryonic kidney cells, and untransformed human fibroblasts as an example. The described protocol can be easily established and it can be adapted to other cell lines or other parameters important for cellular health.

Key words High-content imaging, Multiplex, Machine learning, Phenotypic screening, Cell viability

1 Introduction

In recent years, the use of chemogenomic compound libraries for the identification of biological effects associated with specific targets has gained interest in the scientific community [1–3]. The majority of compounds used in such chemogenomic libraries are well validated in terms of target-compound correlations, but information about their suitability to be used in complex biological systems is often lacking [4]. One way to prevent wrong annotation of phenotypic readouts as a result of the inhibition of specific targets or false positives in scientific research is the prior annotation of chemogenomic compounds, based on their effect on cellular health in different cellular systems. In our laboratory, we use the multiplex assay described here, as a secondary screen for compounds that show a reduced cell growth rate in a primary viability assessment, using a label-free brightfield method described in Chap. 6. There, the compounds are validated based on their

59

calculated growth rate (GR) [5] after 24 h. All "hit" compounds, which are classified as "cytotoxic," when containing a GR value smaller than 0, can be tested further. However, the multiplex method can be used as primary screen as well.

In the multiplex high-content assay, basic cellular functions such as cell viability, mitochondrial health, membrane integrity, and interference of compounds with the cytoskeleton are assessed [6]. It is performed in live-cell mode over a time period of 48 h, but can be easily expanded to longer time points. To avoid a too strong bias of the results, we are routinely testing three different cell lines with diverse cellular morphology: human embryonic kidney cells (HEK293T), osteosarcoma cells (U2OS), and untransformed human fibroblasts (MRC-9). Other adherent cells have been used in this assay as well.

Artificial intelligence (AI) techniques greatly improved image analysis [7] in recent years. The large amount of data that can be generated and the easier accessibility of phenotypic screening libraries have taken data evaluation to the next level. Also, morphological alterations, which are not visible by the human eye, can be detected [8].

Here, for an advanced image analysis, the data analysis is performed using a machine learning-based algorithm approach, where the cells are gated according to a "tree principle," using different characteristics as decision points. For this, we applied the CellPathfinder software from Yokogawa. The machine learning-based algorithm was trained by an experienced cell biologist, based on a set of compounds containing references for every gating step (*see* Subheading 2.2) [6]. After detection of cell body ("Cellbody") and nucleus ("Nucleus"), the cells are gated based on different features in categories connected to 1. compound properties, 2. cell properties, and 3. phenotypic properties. First, Hoechst high-intensity objects are recognized based on the Hoechst33342 channel intensity, to detect autofluorescence or precipitation of a compound. Afterwards, all remaining cells, termed "normal," are further evaluated based on their nuclear morphology and gated in either healthy, pyknosed, or fragmented. The addition of Annexin V (*see* Subheading 2.3) enables the differentiation between mitotic and apoptotic cells. For the phenotypic properties, the healthy cells are gated in three different ways: tubulin effect, increase of mitochondrial mass, and membrane permeabilized. All analysis protocols used in this protocol can be found here as follows: https://doi.org/10.5281/zenodo.6415330.

The protocol described here was validated and tested against a compound set of 230 compounds [6]. It is easy adaptable, but must then be optimized accordingly.

For more detailed information on the establishment and use of the protocol, please see Tjaden et al. [6].

2   Materials

To minimize contamination, all working steps involving live samples should be carried out under sterile conditions. All cell culture reagents, including medium, washing solution and reagents for cell passaging should be stored at 4 °C and be warmed up to 37 °C prior to use. Do not use cells with a higher passage number than 35, to ensure genomic integrity [9]. Compounds of interest, as well as reference compounds, can be prepared as aliquots in advance. They can be stored in a sealed plate or closed vial at −20 °C for a long time period. If used regularly, compounds can be stored at room temperature (RT) for 3–6 month, to avoid freeze–thaw cycles [10]. Chemical integrity, stability, and solubility of the compounds should be validated beforehand (*see* **Note 1** and Chap. 6).

2.1   Cell Culture

The described protocol is optimized and validated to test the adherent cell lines human embryonic kidney cells (HEK293T), osteosarcoma cells (U2OS), and untransformed human fibroblasts (MRC-9) at once. It can be adapted to test just one cell line or different adherent cell lines (*see* **Note 2**).

1. Cell lines: HEK293T (ATCC® CRL-1573™), U2OS (-ATCC®HTB-96™), MRC-9 fibroblasts (ATCC® CCL-2™).

2. Culture medium for HEK293T cells and U2OS cells: Dulbecco's modified Eagle's medium (DMEM) (Gibco) plus L-glutamine (High glucose) supplemented with 10% FBS (Gibco) and penicillin/streptomycin (Gibco).

3. Culture medium for MRC-9 cells: Eagle's Minimum Essential Medium (EMEM) (Gibco) supplemented with 10% FBS (Gibco) and penicillin/streptomycin (Gibco).

4. Fetal bovine serum (FBS).

5. Penicillin (100 U/mL)/streptomycin (0.1 mg/mL) (Gibco).

6. Cell washing solution: Dulbecco's phosphate-buffered saline 1X (DPBS): without calcium or magnesium (Thermo Fisher Scientific).

7. Dissociation reagent: trypsin-EDTA 1X (0.05%) (Thermo Fisher Scientific).

8. Imaging plate: 384-well cell culture microplate, PS, f-bottom, μClear® (Greiner).

2.2   Preparation of Compounds

The protocol described here is optimized to test 135 compounds at two different concentrations (1 μM and 10 μM). However, this can be adapted (*see* **Note 3**). The compounds of interest, as well as reference compounds, should be diluted in a suitable solvent, preferably DMSO, where solubility is anticipated according to the supplier. Precipitated compounds should be excluded before
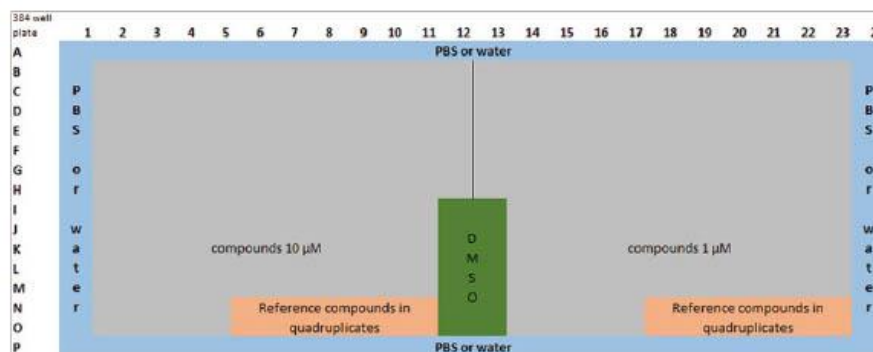
Fig. 1 Example layout of 384-well plate to test 135 compounds at two different concentrations. (This figure is adapted from Tjaden et al. [17])

testing (*see* **Note** 1). Reference compounds with known mode of actions must be added to every experiment, because they serve as training images for the subsequent analysis. As negative control, 10% of the plate should contain wells with 0.1% DMSO (*see* **Note** 4). An example layout of the plate can be found in Fig. 1. For each cell property, at least one reference compound must be tested. As part of the assay development, we tested the following reference compounds [6] and confirmed their suitability to generate a training dataset as follows:

1. Compounds are tested at 1 μM and 10 μM in technical duplicates.

2. Reference compounds should be added in quadruplicates at a final concentration of 10 μM.

   (a) Staurosporine: apoptotic cells, pyknosed nuclei [11].

   (b) Paclitaxel: change of tubulin structure [12].

   (c) Milciclib: increase of mitochondrial mass [13].

   (d) Daunorubicin: fragmented nuclei, apoptotic cells [14].

   (e) Digitonin: permeabilized membrane [15].

   (f) Berzosertib: Hoechst high-intensity objects [16].

3. 10% of the plate should be filled with DMSO 0.1% to serve as negative controls (*see* Fig. 1).

2.3 Cell Staining Dyes

The following dyes were validated against all three cell lines [6]. The concentrations given here should be used, as higher concentrations can affect cell viability and lower concentrations might interfere with the detection power of the analysis software (*see* **Note** 5). Using different fluorescent dyes can lead to bleed-through problems. For the here described protocol, the overlap of

Appendix D

fluorescence emission spectra is neglectable for all dyes but the MitoTracker Red and Annexin V Alexa Fluor 680. However, this overlap does not influence the analysis, since the excitation maxima of these two dyes are well separated during the gating. *See* also [6].

1. 57 μL of 16.23 μM stock of Hoechst33342 (Thermo Fisher Scientific) (final assay concentration: 60 nM).

2. 8 μL of BioTracker™ 488 Green Microtubule Cytoskeleton Dye (EMD Millipore).

3. 12 μL of 100 μM stock of MitoTracker red (Invitrogen) (final assay concentration: 75 nM).

4. 93 μL of Annexin V Alexa Fluor 680 conjugate (Invitrogen).

2.4 Instruments for High-Content Imaging and Parameters

1. CQ1 confocal microscope (Yokogawa) for image acquisition equipped with a CO_2 incubation chamber.

2. Five channels are used for image acquisition: Ex 405 nm/Em 447/60 nm, 500 ms, 50%; Ex 561 nm/Em 617/73 nm, 100 ms, 40%; Ex 488/Em 525/50 nm, 50 ms, 40%; Ex 640 nm/Em 685/40, 50 ms, 20%; brightfield, 300 ms, 100% transmission.

3. Magnification 10×.

4. Focus and Z stacks: 7 Z stacks with a total of 55 μm spacing, focus area for a 384 well plate with a height of 14.4 ± 0.1 mm at -15.5 μm for the lowest (-20.2 μm) and highest focus of 39.7 μm (51.6 μm).

5. Cytomat2C24 incubator (Thermo Fisher Scientific).

2.5 Analysis Software

1. CellPathfinder image analysis software (Yokogawa).

2. Analysis protocols using the CellPathfinder software for all three cell lines can be found here as follows: https://doi.org/10.5281/zenodo.6415330.

3 Methods

3.1 Preparation of Cells for Live-Cell Imaging

1. Culture HEK293T and U2OS over 2 weeks in DMEM plus L-glutamine (high glucose) supplemented with 10% FBS (Gibco) and penicillin/streptomycin (Gibco).

2. Culture MRC-9 fibroblasts in EMEM plus L-glutamine supplemented with 10% FBS (Gibco) and penicillin/streptomycin (Gibco) over the same time period.

3. When cells reach a confluence of approximately 80%, count all cell lines individually using either an automated cell counter or an hemocytometer and assess the cell number and viability (viability above 95% is acceptable).

64   Amelie Tjaden et al.

4. Dilute the cells in 16 mL media to create a cell suspension containing the desired cell number for one 384-well plate. We recommend the following cell count per well (*see* **Note 6**) as follows:

   (a) HEK293T—1500 cells/well.

   (b) U2OS—1500 cells/well.

   (c) MRC-9—1250 cells/well.

5. Add the cell staining dyes described in Subheading 2.3. to the cell suspensions.

6. Prepare a 384-well plate for every cell line (in total three plates).

7. Mix the cells with the dye solution gently and seed 50 μL per well to a 384-well plate (*see* Subheading 2.1). The outer wells should be excluded to avoid evaporation effects (*see* **Note 7**) and can be filled with 100 μL PBS or water (*see* Fig. 1).

8. Remove any air bubbles (*see* **Note 8**).

9. Leave the plate for 30 min at room temperature (RT) to allow reattachment of the cells.

10. Incubate the plate for 18–24 h at 37 °C and 5% CO_2.

3.2 Image Acquisition of Non-treated Cells

1. Switch on the CQ1 microscope system and pre-warm the incubator. Ensure that the environmental controls regulating temperature and CO_2 partial pressure are working.

2. Turn on the laser light and let it heat up for approximately 3 min.

3. The three plates are tested one after another. It is important to keep the order of the different cell lines to minimize plate variations (*see* **Note 9**).

4. Place the first plate prepared in Subheading 3.1. into the CQ1. The other plates should be kept in the incubator until they will be measured.

5. Open the CQ1 software.

6. Create a new protocol for image acquisition.

   (a) Open the "sample" section and select the used 384-well plate with a height of 14.4 ± 0.1 mm. If your plate has not already been implemented in the software, please notify the responsible person for the instrument to implement the plate (*see* **Note 10**).

   (b) Go to the "imaging" section and select the parameters described in Subheading 2.4.

   (c) Field: Set the field in the middle of the well.

Appendix D

(d) Select the blue channel and click on the autofocus "AF" button to automatically search for the best cell area on 3 to 7 different wells. If using the autofocus for your experiment, it can happen that the microscope will sometimes focus on dust and will lose focus (*see* **Note 11**).

(e) Test the set focus again on 3 to 7 different wells to check if your focus is on the correct area. Add this area as standard focus height.

7. Save your protocol.

8. Start measurement of the first plate.

9. Repeat the **steps 4** to 7 for all cell lines tested.

3.3 Image Acquisition of Treated Cells

1. Add the compounds as well as reference compounds to the first plate (*see* Subheading 2.2.).

2. Centrifuge the plate at $100\times g$ for 3 min.

3. Wait 1.5 h and repeat **steps 1** and **2** for the second plate (*see* **Note 9**).

4. Wait again 1.5 h and repeat **steps 1** and **2** for the third plate.

5. Incubate the plates for 12 h after treatment at 37 °C and 5% CO_2. Other time points for image acquisition can be used. The incubation time must be adapted accordingly (*see* **Note 9**).

6. Place the first plate into the CQ1.

7. Open the measurement protocol that was used for the image acquisition of non-treated cells prepared in Subheading 3.2.

8. Measure the plate.

9. After measuring, incubate the plate again for 12 h (*see* **Note 9**).

10. Repeat **steps 6** to **8** again after 24 h and 48 h after compound treatment (*see* Subheading 3.3, **step 1**).

11. Repeat steps 6 to 10 with the other two plates, respectively.

12. A scheme of the workflow can be found in Fig. 2.

### 3.4 Data Analysis Using CellPathfinder Software

1. Open CellPathfinder software.

2. Convert CQ1 images as instructed by the software.

3. Open the 24 h data of one cell line first to create an analysis protocol for this cell line (*see* **Note 12**) or use the provided protocols (*see* Subheading 2.4, **step 2**). The following steps should be performed accordingly with the other cell lines.

4. The following channels are used for the analysis as follows:

(a) CH1: Hoechst 33342 (DNA detection): Ex 405 nm/Em 447/60 nm, MaxIP.

(b) CH2: BioTracker™ 488 Green Microtubule Cytoskeleton Dye (tubulin stain): Ex 488/Em 525/50 nm, MaxIP.

148

66   Amelie Tjaden et al.
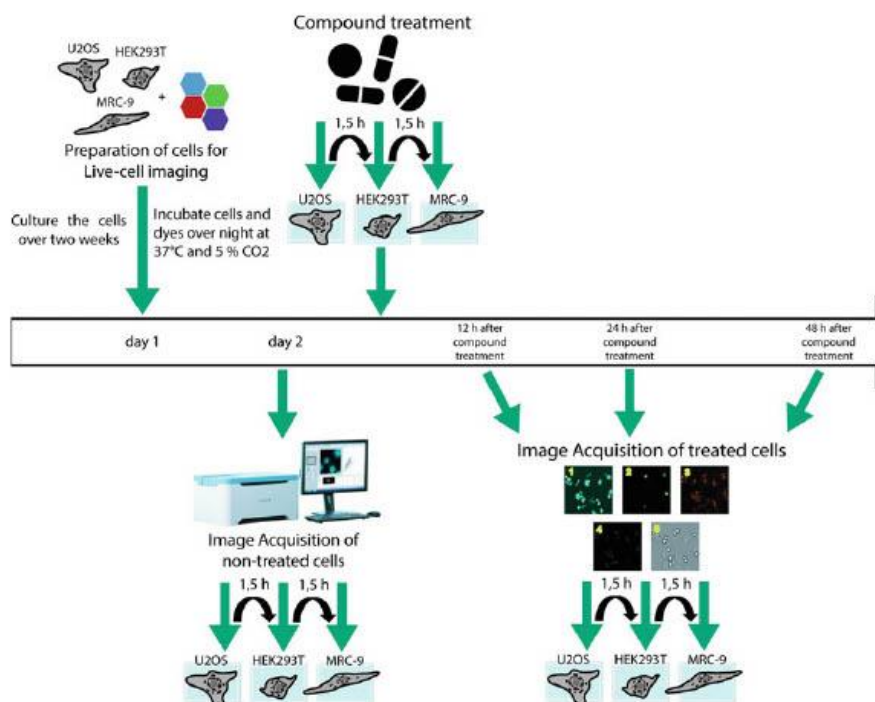
Fig. 2 Scheme of workflow for three cell lines

   (c) CH3: MitoTracker red (mitochondrial mass detection): Ex 561 nm/Em 617/73 nm, MaxIP.

   (d) CH4: Annexin V (apoptosis marker): Ex 640 nm/Em 685/40, MaxIP.

   (e) CH5: brightfield, DCP mode "fluor-" type.

5. For object detection, create the two objects "Cellbody" and "Nucleus" (*see* Fig. 3).

   (a) "Nucleus": Ex 405 nm/Em 447/60 nm, Finder: Nuclear, Recognition: Advance.

   (b) "Cellbody": brightfield, Finder: Cell, Recognition: Advance.

6. Define the object "Nucleus" to be included in the "Cellbody" (*see* Fig. 3).

7. After object detection, the gating is performed using the machine learning (ML) based function implemented in the software.
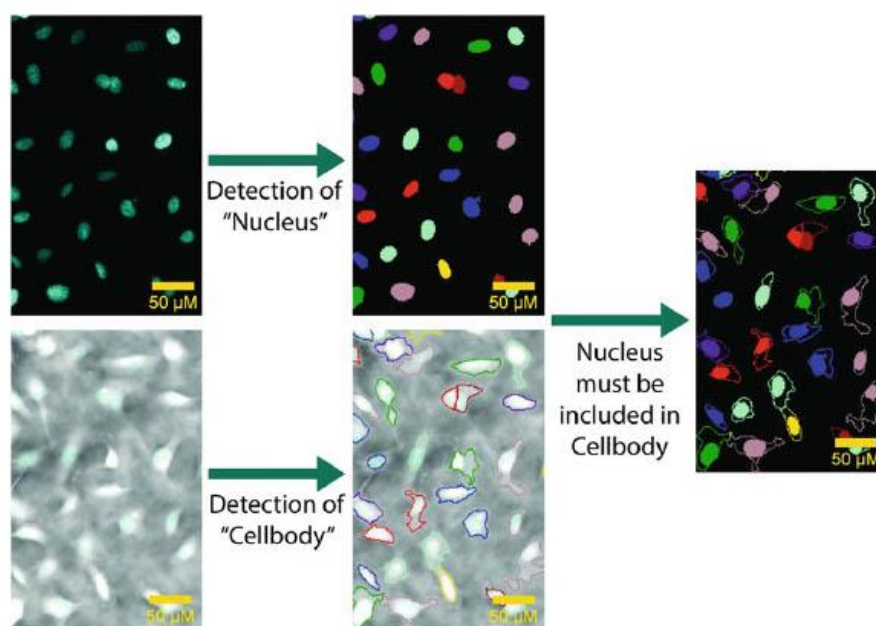
Fig. 3 Object detection. The two objects "Cellbody" and "Nucleus" should be detected by the software algorithm. The Nucleus should be defined as included in the Cellbody

    (a) Use DMSO 0.1% wells as negative controls (*see* **Note 4**).

    (b) Use the reference compounds (*see* Subheading 2.2) to train the ML algorithm (*see* **Note 13**).

    (c) To validate your algorithm, check other wells containing the reference compounds (*see* **Note 13**).

8. The gating hierarchy can be found in Fig. 4.

9. All features used for training the machine learning algorithm are shown in Table 1.

10. Select 10 to 15 cells for every gating step. A detailed explanation of how to train the machine learning-based algorithm can be found in the CellPathfinder user information as well as in the recently published protocol by Tjaden et al. [17].

11. Save the analysis protocol.

12. Analyze the data from all time points using the same analysis.

13. Repeat the **steps 3** to **12** for the other cell lines.

68   Amelie Tjaden et al.
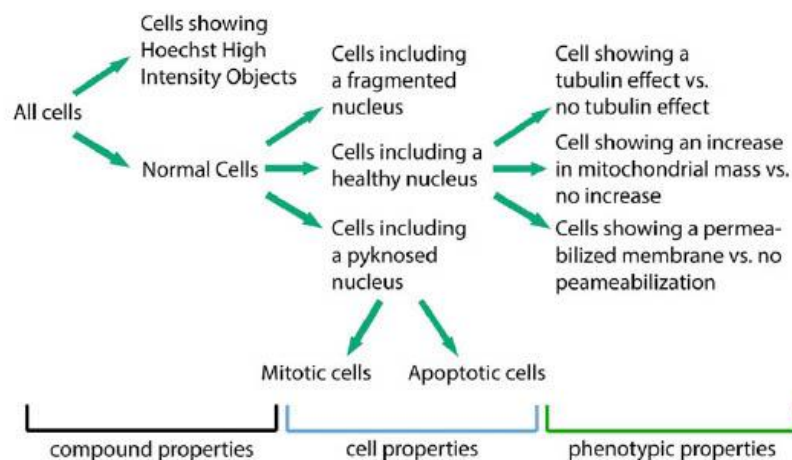
Fig. 4 Gating hierarchy. The gating is performed employing a tree principle. First, the compounds properties (Hoechst high-intensity objects or normal cells) are defined. Afterward, the cells are gated based on their nuclear morphology in healthy, pyknosed, or fragmented. Cells containing a pyknosed nucleus are further gated in apoptotic or mitotic cells. Lastly, phenotypic properties are considered to divide cells into the categories tubulin effect/no tubulin effect, mitochondrial mass increased/no increase, and membrane permeabilized/no permeabilization

3.5 Data Evaluation

1. After analyzing all cell lines for all time points, results are evaluated.

2. You can save the images under the "view" tab.

3. Growth rate can be calculated as shown in Chap. 6.

4. For biological data, at least two biological replicates should be performed.

5. The cell count of all gating steps can be used to calculate the ratios of the different properties (*see* **Note 14**) (*see* Fig. 5).

6. All data should always be compared with the screened positive and negative controls.

7. For a basic annotation, we defined the threshold of 50% to define a compound as "hit" compound. Compounds that are "hits" in more than one cell line should be evaluated further. The phenotypic evaluation can be optimized further if needed for more in-depth annotation regarding mitochondrial and tubulin features.

Table 1
Gating features used for machine learning algorithm

Gating type	Cell region	Features	Gating type	Cell region	Features
Hoechst high-intensity objects vs. normal cells	Cellbody	Mean intensity CH1	Healthy, pyknosed, or fragmented nuclei	Nucleus	Total intensity CH1
	Cellbody	Min intensity CH1		Nucleus	Total hole CH1
	Cellbody	Max intensity CH1		Nucleus	Total valley CH1
	Nucleus	Max intensity CH1		Nucleus	Total edge CH1
	Nucleus	Nuc_cell_area		Nucleus	Total saddle CH1
	Nucleus	Nuc area/cellbody area		Nucleus	Mean hole CH1
Mitotic vs. apoptotic cells	Cellbody	Total intensity CH4		Nucleus	Mean ridge CH1
	Cellbody	Max intensity CH4		Nucleus	Mean valley CH1
	Cellbody	Mean peak CH4		Nucleus	Mean edge CH1
	Cellbody	Mean hole CH4		Nucleus	Mean saddle CH1
	Cellbody	Mean ridge CH4	Tubulin effect vs. no tubulin effect	Cellbody	Total intensity CH2
	Cellbody	Mean valley CH4		Cellbody	Mean intensity CH2
	Cellbody	Mean saddle CH4		Cellbody	Max intensity CH2
	Nucleus	Total hole CH1		Cellbody	Mean peak CH2
	Nucleus	Mean peak CH1		Cellbody	Mean ridge CH2
	Nucleus	Mean hole CH1		Cellbody	Mean edge CH2
	Nucleus	Mean ridge CH1		Cellbody	Total intensity CH2
	Nucleus	Mean valley CH1	Membrane permeabilized vs. normal	Cellbody	Total intensity CH4
Mitochondrial mass increased vs. normal	Cellbody	Total intensity CH3		Cellbody	Mean intensity CH2
	Cellbody	Mean intensity CH3		Cellbody	Mean intensity CH3
	Cellbody	Max intensity CH3		Cellbody	Max intensity CH2
	Cellbody	Total peak CH3		Cellbody	Total intensity CH2
	Cellbody	Mean saddle CH3		Cellbody	Mean peak CH2
	Cellbody	Mean peak CH3		Cellbody	Mean peak CH3
	Cellbody	Mean hole CH3		Cellbody	Mean peak CH5
	Cellbody	Mean ridge CH3		Cellbody	Mean edge CH2
	Cellbody	Mean valley CH3		Cellbody	Mean hole CH2
				Cellbody	Mean hole CH3
				Cellbody	Mean ridge CH2
				Cellbody	Mean ridge CH3
				Cellbody	Mean ridge CH5
				Cellbody	Mean valley CH2
				Cellbody	Mean valley CH3
				Nucleus	Compactness

70   Amelie Tjaden et al.
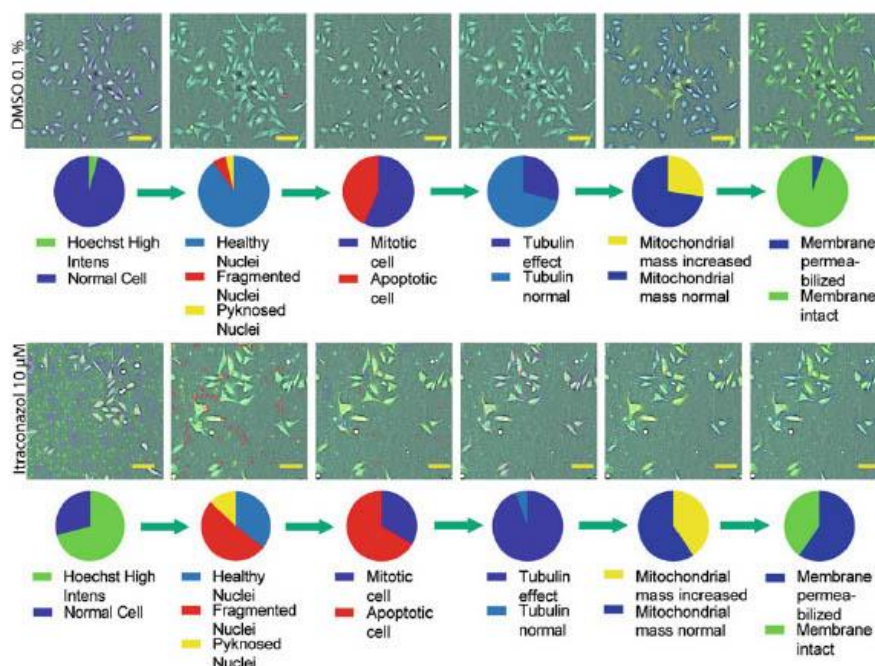
Fig. 5 Processed image of stained (blue: DNA/nuclei, green: microtubule, red: mitochondria content, magenta: Annexin V apoptosis marker) U2OS cells after 24 h of exposure to Itraconazol (10 μM) and 0.1% DMSO. U2OS cells are gated into different categories using the CellPathfinder analysis. Pie charts show the ratios of the different gating steps. (This figure is adapted from Tjaden et al. [17])

4   Notes

1. All compounds should be checked by eye if precipitation happened during storing. Precipitated compounds look cloudy in the dilution or on the plate. If precipitation happened, we recommend to freshly dissolve the compound or use a different solvent, e.g., water. A more professional check can be performed as described in Chaps 4 and 6.

2. If other cell lines than the once described here are used, the protocol must be adapted accordingly. The dye concentration, incubation time, and cell seeding density should be tested using a reference compound set. For more information on how to perform this, please see Tjaden et al. [6].

3. Other compound concentrations can be tested. If more than two are tested, the plate layout (*see* Fig. 1) must be adapted accordingly.

4. The reference compounds are used to train the machine learning-based algorithm. They should be added in quadruplicates to validate the analysis in different reference compound wells. Other compounds can also be used, when they have a similar mode of action. The machine learning-based algorithm has to be trained accordingly. The negative control is important to normalize your results and to evaluate the ratios calculated based on the gating steps in comparison to the controls.

5. All dyes used in the protocol described here were tested at different concentrations over time in all three cell lines. The concentration used was validated as the best concentration for a robust fluorescent readout without having an impact on cell viability. If other dyes or concentrations are used, the protocol must be adapted accordingly.

6. The cell count per well was tested in a pre-experiment to detect the best cell seeding density for testing the cells over 48 h. For more information about how to select the best cell seeding concentration, *see* Tjaden et al. [17].

7. Evaporation of the media and "edge-effects" can be minimized when filling the outer wells with a buffer of choice, e.g., DPBS or water in a volume of 100 µL per well.

8. Air bubbles can interfere with cell growth or later image acquisition. They should be removed using a "debubbler" or something similar. The debubbler can be a nozzle inserted into an empty bottle containing, e.g., ethanol. The nozzle should not touch the liquid, so that air can be expelled to remove any bubbles.

9. For image acquisition of one 384-well plate, the CQ1 confocal microscope approximately needs 1.5 h. This time difference should be kept in mind when pipetting the compounds. The plate order when testing different plates should always be the same. The time points of image acquisition can be changed if desired. For three plates the time points in between measurements must be more than 4.5 h.

10. If another plate is used than implemented in the CQ1, focus might be lost during the measurement. We recommend to add the plate information to the CQ1 software before measurement. This can be done by the Yokogawa support team.

11. When using autofocus for the whole measurement, the CQ1 sometimes focuses on dust on the lid instead of cells and loses the best focus area. We recommend to use the autofocus to search for the best cell area and enter this focus area as a standard for the whole plate.

72 Amelie Tjaden et al.

12. The cell lines differ in their morphology and fluorescence intensity levels. Therefore, individual analysis protocols should be created for each cell line.

13. For the machine learning based algorithm, approximately 10 to 15 example cells of the reference compounds should be selected. To validate the analysis, other wells of the same reference compounds should be used.

14. Keeping the tree principle in mind, cell count ratios of the different gating steps can be calculated by dividing the cell count of a gating step by the cell count of next higher level. Cell count ratios can then be compared with control wells.

Acknowledgement

References

1. Bredel M, Jacoby E (2004) Chemogenomics: an emerging strategy for rapid target and drug discovery. Nat Rev Genet 5(4):262–275

2. Caron PR et al (2001) Chemogenomic approaches to drug discovery. Curr Opin Chem Biol 5(4):464–470

3. Wells CI et al (2021) The kinase Chemogenomic set (KCGS): an Open Science resource for kinase vulnerability identification. Int J Mol Sci 22(2):566

4. Gerry CJ et al (2016) Real-time biological annotation of synthetic compounds. J Am Chem Soc 138(28):8920–8927

5. Hafner M et al (2016) Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. Nat Methods 13(6):521–527

6. Tjaden A et al (2022) Image-based annotation of chemogenomic libraries for phenotypic screening. Molecules 27(4):1439

7. Boyd J, Fennell M, Carpenter A (2020) Harnessing the power of microscopy images to accelerate drug discovery: what are the possibilities? Expert Opin Drug Discovery 15(6): 639–642

8. Ziegler S, Sievers S, Waldmann H (2021) Morphological profiling of small molecules. Cell Chem Bio 28(3):300–319

9. Hughes P et al (2007) The costs of using unauthenticated, over-passaged cell lines: how much more data do we need? BioTechniques 43(5):575–586

10. Kozikowski BA et al (2003) The effect of room-temperature storage on the stability of compounds in DMSO. J Biomol Screen 8(2): 205–209

11. Bruno S et al (1992) Different effects of staurosporine, an inhibitor of protein kinases, on the cell cycle and chromatin structure of normal and leukemic lymphocytes. Cancer Res 52(2): 470–473

12. Wang TH, Wang HS, Soong YK (2000) Paclitaxel-induced cell death: where the cell cycle and apoptosis come together. Cancer 88(11):2619–2628

13. Sanchez-Martinez C et al (2015) Cyclin dependent kinase (CDK) inhibitors as anticancer drugs. Bioorg Med Chem Lett 25(17): 3420–3435

14. Al-Aamri HM et al (2019) Time dependent response of daunorubicin on cytotoxicity, cell cycle and DNA repair in acute lymphoblastic leukaemia. BMC Cancer 19(1):179

15. Styrt B, Johnson PC, Klempner MS (1985) Differential lysis of plasma membranes and granules of human neutrophils by digitonin. Tissue Cell 17(6):793–800

16. Fokas E et al (2012) Targeting ATR in vivo using the novel inhibitor VE-822 results in selective sensitization of pancreatic tumors to radiation. Cell Death Dis 3(12):e441–e441

17. Tjaden A et al (2022) High-content live-cell multiplex screen for chemogenomic compound annotation based on nuclear morphology. STAR Protocols 3(4):101791

Appendix E: Deep Annotation of Donated Chemical Probes (Dcp) in Organotypic Human Liver Cultures and Patient-Derived Organoids from Tumor and Normal Colorectum

Reprinted with permission from: Tredup, Claudia, Benardina Ndreshkjana, Natalie S. Schneider, Amelie Menge, Aurino M. Kemas, Sonia Youhanna, Volker M. Lauschke, Benedict-Tilman Berger, Andreas Krämer, Lena M. Berger, Sandra Röhm, Stefan Knapp, Henner F. Farin, and Susanne Müller. "Deep Annotation of Donated Chemical Probes (Dcp) in Organotypic Human Liver Cultures and Patient-Derived Organoids from Tumor and Normal Colorectum." ACS Chemical Biology (2023/03/21 2023). https://dx.doi.org/10.1021/acschembio.2c00877. 2023 The Authors. Published by American Chemical Society. This publication is licensed under CC-BY-NC-ND 4.0.

Notice that further permission related to the material excerpted should be directed to the publisher.

Contribution Menge, A.: Viability assessment using Multiplex high-content live-cell assay, contribution to study design, data evaluation, figure preparation, manuscript editing.

ACS chemical biology

Deep Annotation of Donated Chemical Probes (DCP) in Organotypic Human Liver Cultures and Patient-Derived Organoids from Tumor and Normal Colorectum

Claudia Tredup,*[¶] Benardina Ndreshkjana,[¶] Natalie S. Schneider, Amelie Tjaden, Aurino M. Kemas, Sonia Youhanna, Volker M. Lauschke, Benedict-Tilman Berger, Andreas Krämer, Lena M. Berger, Sandra Röhm, Stefan Knapp, Henner F. Farin, and Susanne Müller*
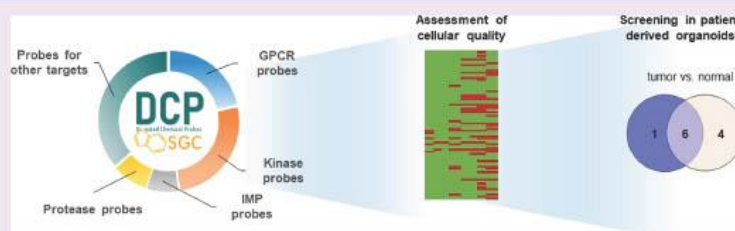
Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Well-characterized small molecules are essential tools for studying the biology and therapeutic relevance of a target protein. However, many compounds reported in the literature and routinely studied in biomedical research lack the potency and selectivity required for mechanistic cellular studies on the function of a given protein. Furthermore, commercially available compounds often do not include useful tools developed by industry as part of their research and development efforts, as they frequently remain proprietary. The freely available donated chemical probe (DCP) library, fueled by generous donations of compounds from industry and academia, enables easy access to a steadily growing collection of these valuable and well-characterized tools. Here, we provide a systematic description of the current DCP library collection and their associated comprehensive characterization data, including a variety of *in vitro* and cellular assays. Of note, we characterized the set in relevant human primary models by employing hepatotoxicity screening in primary human liver spheroids and viability screening in patient-derived colorectal cancer organoids and matched normal-adjacent epithelium. Taken together, the DCP library represents a well-annotated, openly available collection of tool compounds for studying a wide range of targets, including kinases, G-protein-coupled receptors, and ion channels. As such, it represents a unique resource for the biomedical research community.

■ INTRODUCTION

High-quality chemical tool compounds are widely accepted as one of the most versatile way to modulate and study the function of a protein. However, many compounds reported in the literature and routinely used in laboratory practice either lack selectivity, stability, and potency needed to achieve exact perturbation of a single target or they are used under conditions that favor off-target and pleiotropic effects.[1−5]

One of the most notable efforts in this area was the establishment of standard criteria for evaluation of compounds as chemical tools.[2] In general, a compound can be considered to be a chemical probe if it fulfills the following criteria: (a) exhibits potency on the intended target <100 nM (IC$_{50}$ or K_D); (b) exhibits selectivity within the target family >30-fold; (c) has undergone extensive off-target profiling outside the target family; (d) exhibits cellular on-target activity <1 μM (IC$_{50}$ or EC$_{50}$); and (e) is tested in comparison with 100-fold less potent control compound of the same chemical scaffold (paired control). Ideally, the compounds should not contain reactive chemical elements known as PAINS elements.[6]

As a consequence, a growing number of resources have been made available to scientists, helping to find the most suitable chemical tools for their research. Depending on the scientific

ACS Publications

822

Table 1. DCP Set of 97 Chemical Probes (As of June 2022)

target family	target	probe	mode of action	control
kinase	ACVR1B,TGFBR1	TP-008	inhibitor	AI11
	AKT1, AKT2	BAY1125976	allosteric Inh.	BAY-940
	AKT1, AKT2	Borussertib	covalent-allosteric Inh.	RL2578
	CLK1/2/3/4	T3-CLK	inhibitor	T3-CLK-N
	DDR1/2, MAPK11, MAPK14	SR-302	inhibitor	SR-301
	IKBKB	BI 605906	inhibitor	BI-5026
	IRAK4	GNE-2256	inhibitor	GNE-6689
	LRRK2	MLi-2	inhibitor	MLi-2-NC
	MAPK1/3	ERKi	inhibitor	ERKi-NC
	MAPK7	BAY-885	inhibitor	BAY-693
	MAPK14	FS-694	type 1.5 Inh.	FM-743
	MAPK14	Skepinone-L	type 1 Inh.	FM-743
	MAPK14	SR-318	type 2 Inh.	SR-321
	MET	BAY-474	inhibitor	BAY-827
	PIP4K2A	BAY-091	inhibitor	BAY-0361
	PRKAA1, RPS6KA1	BAY-3827	inhibitor	BAY-974
	PTK2,PTK2B	PF-04554878	type 1 Inh.	PF-00911705
	RIPK1	TP-030-1	inhibitor	TP-030n
	RIPK1	TP-030-2	inhibitor	TP-030n
	ROCK1/2	BAY-549	inhibitor	BAY-4900
	SYK	MRL-SYKi	inhibitor	MRL-SYKi-NC
	TBK1, IKBKE	BAY-985	inhibitor	BAY-440
	TIE1,TEK, DDR1/2	BAY-826	inhibitor	BAY-309
GPCR	ADRA1D	(R)-9s	antagonist	(S)-9s
	CCR1	BI 639667	antagonist	BI-9307
	CCR1	BAY-3153	antagonist	BAY-173
	CHRM1	MSD-M1PAM	PAM	MSD-M1PAM-NC
	CNR1	MRL-650	inverse agonist	MRL-CB1-NC
	DRD4	ABT-724	agonist	A-769
	DRD4	UCSF924	agonist	UCSF924NC
	EDNRA	ABT-546	antagonist	A-545
	EDNRB	A-192621	antagonist	A-1806262
	F2R	BAY-386	antagonist	BAY-448
	GNRHR	BAY-784	antagonist	BAY-786
	GPR52	TP-024	agonist	TP-024n
	GPR58	Ogerin	PAM	ZINC32547799
	KISS1R	KISS1-305	agonist	KISS1-543
	LHCGR	BAY-899	allosteric antagonist	BAY-897
	MRGPRX2	(R)-ZINC-3573	agonist	(S)-ZINC-3573
	P2RY14	PPTN	antagonist	PPTN-NC
	PTGDR2	CRTH2 antagonist	antagonist	CRTH2 negative control
	PTGER2	PF-04418948	antagonist	PF-04475866
	PTGFR	BAY-6672	antagonist	BAY-403
ion channel	CFTR	A-1596586	corrector	A-1596584
	P2RX4	BAY-1797	inhibitor	BAY-207
	SLC9A1	BI-9627	antagonist	BI-0054
	TRPA1	A-079	antagonist	A-226
	TRPA1	BAY-390	inhibitor	BAY-9897
	TRPM8	PF-05105679	antagonist	PF-05257137
epigenetics	BRD2/3/4, BRDT(BD1)	GSK778	inhibitor	no
		GSK789	inhibitor	GSK791
	BRD2/3/4, BRDT(BD2)	GSK046	inhibitor	no
		GSK620	inhibitor	no
		GSK973	inhibitor	GSK943
	EP300,CREBBP	A-485	inhibitor	A-486
	KAT6A, KAT6B	WM-1119	inhibitor	WM-2474
other enzymes	ALOX5AP	BI 665915	inhibitor	BI-0153
	BCAT1/2	BAY-069	inhibitor	BAY-771
	CMA1	BI-1942	inhibitor	no
	Complex I	BAY-179	inhibitor	BAY-070
	CYP11B2	MSD-CYP11B2	inhibitor	MSD-CYP11B2 negative control

Table 1. continued

	DHODH	IPP/CNRS-A017	inhibitor	IPP/CNRS-A019
	ELANE	BAY-678	inhibitor	BAY-677
	ENPP2	BI-2545	inhibitor	BI
	EPHX2	BI-1935	inhibitor	BI-2049
	FAAH	PF-04457845	covalent Inh.	PF-04875474
	FASN	BI 99179	inhibitor	BI 99990
	FNTB	ABT-100	inhibitor	A-108
	GLS	A-446	inhibitor	A-426
	γ secretase complex	GSM1	modulator	GSM-NC
		MRK-560	inhibitor	GSI-NC
	HCV NS3	BI-1230	inhibitor	BI-1675
	HCV NS5B	BI 207127	inhibitor	BI-7656
	HIV NNRT	BI-2540	inhibitor	BI-2439
	MALT1	NVS-MALT1	allosteric Inh.	NVS-MALT1-C
	METAP2	TP-004	inhibitor	TPn-004
	MGAT2	TP-020	inhibitor	TP-020n
	MIF	BTZO-1	ARE activator	BTZO-4
	MMP12	BAY-7598	inhibitor	BAY-694
	MMP13	T-26c	inhibitor	T-26f
	MMP13	BI-4394	inhibitor	BI-4395
	NUDT1	BAY-707	inhibitor	BAY-604
	OGA	JNJ-65355394	agonist or Inh.	JNJ-73924149
	OGA	TP-040	inhibitor	TP-040n
	PDE10A	JNJ-42396302	inhibitor	JNJ-40573663
	PDE10A	THPP-1	inhibitor	THPP-1-NC
	SLC2A1	BAY-876	inhibitor	BAY-588
	SOS1	BAY-293	inhibitor	BAY-294
	UCHL1	8RK64	covalent Inh.	JYQ88
	VNN1	PFI-653	inhibitor	PFI-653-N
other targets	BCL2	A-1211212	inhibitor	A-1210227
	BCL2L1	A-1155463	inhibitor	A-1107969
	BCL6	TP-021	inhibitor	TP-021n
	ITGAL	BI-1950	inhibitor	BI-9446
	NR3C1	BI 653048	agonist	BI-3047
	RORC	JNJ-54119936	inverse agonist	JNJ-53721590
	SLC16A3	MSC-4381	inhibitor	MSC-0516

question, researchers may seek for compounds to interrogate the cellular function of a protein, to study the target of a chemical probe in a disease phenotype, or to look for a starting point for further optimization of a compound, e.g., as a degrader molecule.[7] Most of the available search tools cover a large number of inhibitors and chemical modulators and provide the user with the opportunity to choose the best compound for their specific purpose.[8] Databases such as ChEMBL (https://www.ebi.ac.uk/chembl/), binding DB (https://www.bindingdb.org), canSAR (https://cansarblack.icr.ac.uk), ProbeMiner (https://probeminer.icr.ac.uk), and Probes and Drugs (https://www.probes-drugs.org) are addressed to the chemical biology community, offering user-friendly access to data associated with specific compounds.[8,9] However, the compounds included in these databases are not limited to high-quality, well-validated chemical probes and require user appreciation for the chemical probe criteria outlined above. On the other hand, the Chemical Probes Portal (https://www.chemicalprobes.org/)[10] is an expert curated database, where each compound is reviewed by a panel of expert chemical biologists. Therefore, although limited in scope, the Portal provides expert opinions and guidelines on probe use to help researchers navigate the complex process of finding the right tool for their system and their question.

Once the user has identified the best tool compounds to use, most of them can be obtained from commercial vendors. Recently, few noncommercial resources have emerged, offering access to compounds. Among those are opnMe (https://opnme.com/molecules),[11] which offers more than 70 compounds from a variety of research areas, originating from legacy and ongoing projects of Boehringer Ingelheim[12] and the laboratory of Nathanael Gray, where inhibitors, mostly for protein kinases and several PROTAC molecules, are accessible together with the respective profiling data (https://graylab.dana-farber.org/probes.html). Additionally, the Structural Genomics Consortium (SGC) provides chemical probes for epigenetic targets, protein kinases, and E3 ligases (https://www.thesgc.org/chemical-probes) as well as a Kinase Chemo-genomic Set (KCGS)[13] (https://www.sgc-unc.org/kcgs). A chemogenomic set is also available via the Enabling & Unlocking Biology in the OPEN (EUbOPEN) IMI consortium (https://www.eubopen.org/chemogenomics/chemogenomic-set).

A unique set, in terms of the breadth of targets covered by chemical probes is the donated chemical probes (DCP) set.[14] The project was initiated in 2017 by several pharmaceutical companies (AbbVie, Bayer AG, Boehringer Ingelheim, Janssen, MSD, Pfizer, and Takeda), which entered into a precompe-
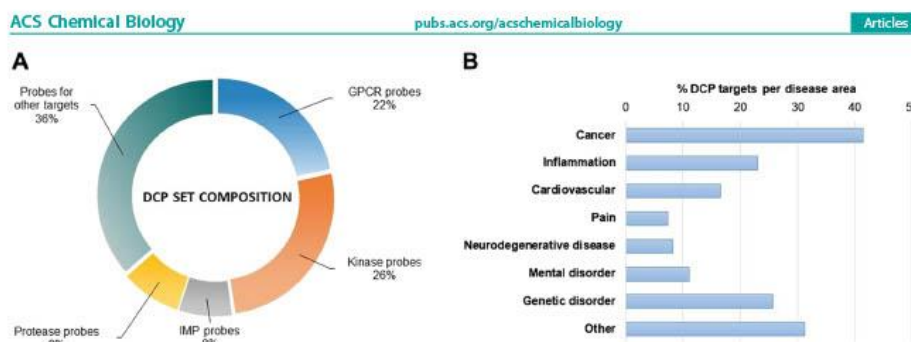
Figure 1. Overview of the DCP set of 97 chemical probes. (A) Percentage of donated chemical probes per target category and (B) implication of 100 DCP targets in different diseases, based on PubMed search. Please note that one target can be implicated in more than one disease area. IMP = integral membrane proteins, GPCR = G-protein-coupled receptors.

titive collaboration to make a large number of high-quality chemical probes available to the public, including their characterization data, inactive control compounds, and recommendations on their use (https://www.sgc-ffm.uni-frankfurt.de//). The initial goal was to assemble a set of 70 chemical probe compounds with their respective controls, perform further characterization in a variety of assays, and make the data as well as the actual compounds accessible to the community in a streamlined manner without high contractual hurdles. The abovementioned guidelines and criteria for chemical probes are evaluated for each compound by an internal and external independent committee, in a two-tier scientific review process before the compounds are included in the set (https://www.sgc-ffm.uni-frankfurt.de/#!donateview). Over the course of the project, not only inhibitors but also activators and positive allosteric modulators (PAM) have been added. Nevertheless, the set represents only a starting point and additional target areas need to be covered by high-quality chemical probes in addition to the proteins represented in the set, which are still biased toward more traditional disease areas such as oncology and inflammation. Hence, with the initial publication,[14] a call went out to the community to help in this endeavor and donate high-quality compounds to the set. Meanwhile, 23 contributors from industry and academia have participated in this unique project and provided not only characterization data but also physical compounds to be distributed among the scientific community. As expected, in some cases, additional profiling data needed to be generated to meet the chemical probe criteria. Data have been provided by either the SGC or resources outside the SGC and we are grateful for the support of the Division of Pre-Clinical Innovation of the National Center for Advancing Translational Sciences (NCATS; https://ncats.nih.gov/) and the National Institute of Mental Health (NIMH) Psychoactive Drug Screening Program (https://pdspdb.unc.edu/pdspWeb/), housed at the laboratory of Dr. Bryan Roth at the University of North Carolina at Chapel Hill. Here, we describe the current composition of this unique resource and the type of data donated and generated over the last 5 years, with the aim to encourage the use of the highly validated compounds as well as donations to complement the set covering additional target proteins.

■ RESULTS AND DISCUSSION

DCP Database and Website. High-quality chemical tools, generated by the pharmaceutical industry often remain undisclosed and unavailable to the research community, thereby impeding scientific progress. In contrast, DCP compounds and all associated profiling data are in the public domain and available through the DCP website (https://www.sgc-ffm.uni-frankfurt.de/), free of charge to all researchers worldwide without any restrictions or need for user registration.

The DCP website was created based on FAIR[15] (Findable, Accessible, Interoperable, and Reusable) data principles, thus allowing user-friendly access to probes and the associated data. As of June 2022, the DCP website provides a curated and accurate source of information for the set of 97 chemical probes (Table 1), with a DOI number for each probe and paired control. Through collaboration with ChEMBL and the Chemical Probes Portal, all compounds possess a unique ChEMBL compound ID and a link to the Chemical Probes Portal to the expert review. The website was developed as a source for different users, such as chemists, biologists, or clinical researchers, allowing a probe, target, or target family search as well as providing a list of the probe set on the overview page (https://www.sgc-ffm.uni-frankfurt.de/#!probes). Powered by a PostgreSQL database (https://www.postgresql.org/about/), detailed potency and selectivity data, advice for the appropriate use in cellular and in *in vivo* models together with assay descriptions, phenotypic assay characteristics, and compound-related structural information are provided. Most data can be downloaded in a database-readable format either for a single compound or as a batch for the whole set. In addition, structure (SD) files as well as analytical data (^{1}H-NMR, a ^{13}C-NMR, HPLC data, measured in-house) and structural properties such as redox activity data[16] are downloadable from the website. Furthermore, a flyer with key data of all DCPs as well as the appropriate concentration to use the compounds is available on the website.

In addition to providing experimental data, the DCP program makes all compounds openly available through a nonbureaucratic and simple distribution process, which involves the supply of individual probes or probe sets under a simple web-accessible Open Science Trust Agreement
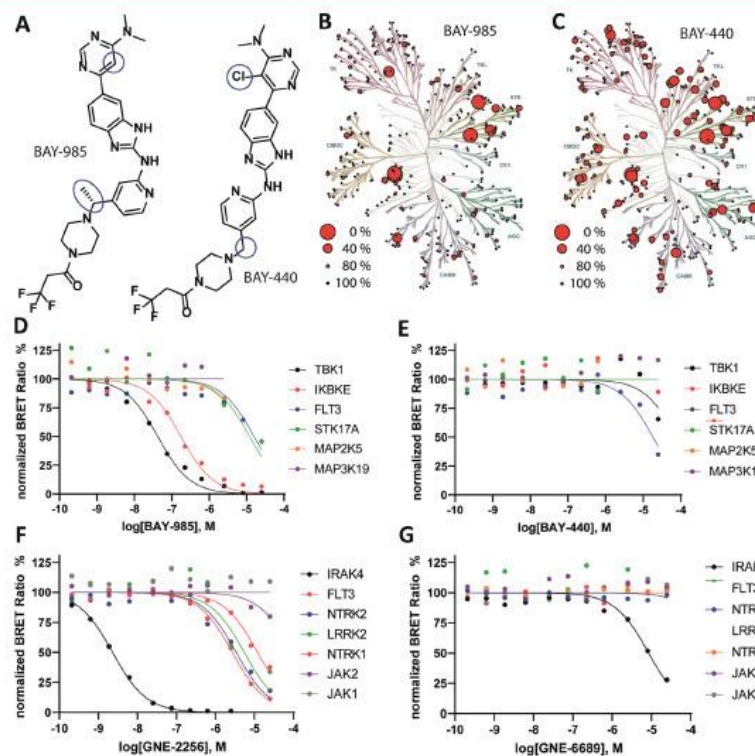
Figure 2. Off-target profiling of the IKBKE/TBK1 probe BAY-985 and IRAK4 probe GNE-2256 showed no considerable off-targets within a 30-fold selectivity range to the probe potency in cells. In vitro data translated well into living cells. (A) Chemical structures of the chemical probe BAY-985 and negative control BAY-440. DiscoverX selectivity assessment for BAY-985 (B) and BAY-440 (C) at 100 nM. Illustration reproduced courtesy of Cell Signaling Technology, Inc. (www.cellsignal.com). The legend is indicated in the figure. NanoBRET in cellulo profiling for BAY-985 (D), BAY-440 (E), GNE-2256 (F), and GNE-6689 (G).

(http://www.thesgc.org/click-trust) as well as through trusted commercial vendors. Since the start of the program, the research community has used the DCP set for a wide variety of applications. More than 12,000 individual compounds and more than 100 sets have been shared with researchers around the world in 100 countries and 28 organizations.

The DCP set covers a broad array of targets from different protein families (Figure 1A and Table 1), which are relevant for a number of disease indications, such as cancer, inflammation, cardiovascular and neurodegenerative diseases, genetic disorders, and other diseases (Figure 1B). Therefore, the set represents a useful tool for target identification and target confirmation in a variety of cellular applications.

Characterization of DCPs. Lack of characterization and annotation of a chemical compound can result in misleading or incorrect conclusions when it is used as a mechanistic tool in biomedical research.[8,17] Therefore, all compounds included in the DCP set are accompanied by characterization data and a negative control compound to minimize this risk. The negative

control compound—a close chemical analogue of the probe with similar physicochemical properties—is inactive or significantly less active on the target. We implemented a 100-fold (or more) decrease in activity as a criterion for the control as compared to the probe compound (Table 1). The negative control compound thus is thought to bind to any potential off-targets, but not to the intended target of the probe. For mechanistic studies or target validation, it is recommended to use these negative controls in parallel to the probe to confirm that the phenotype observed is due to modulation of the intended target and not some off-target activity. Nevertheless, in some cases, the difference between probe compound and matched control compound may also result in the loss of binding of off-targets. Therefore, use of an orthogonal chemical probe with a different scaffold is desirable, as different scaffolds are expected to have different off-target binding profiles and activities.[18] If available, recommendations for chemically unrelated chemical probes for the same target protein are provided on the DCP website. For example, the
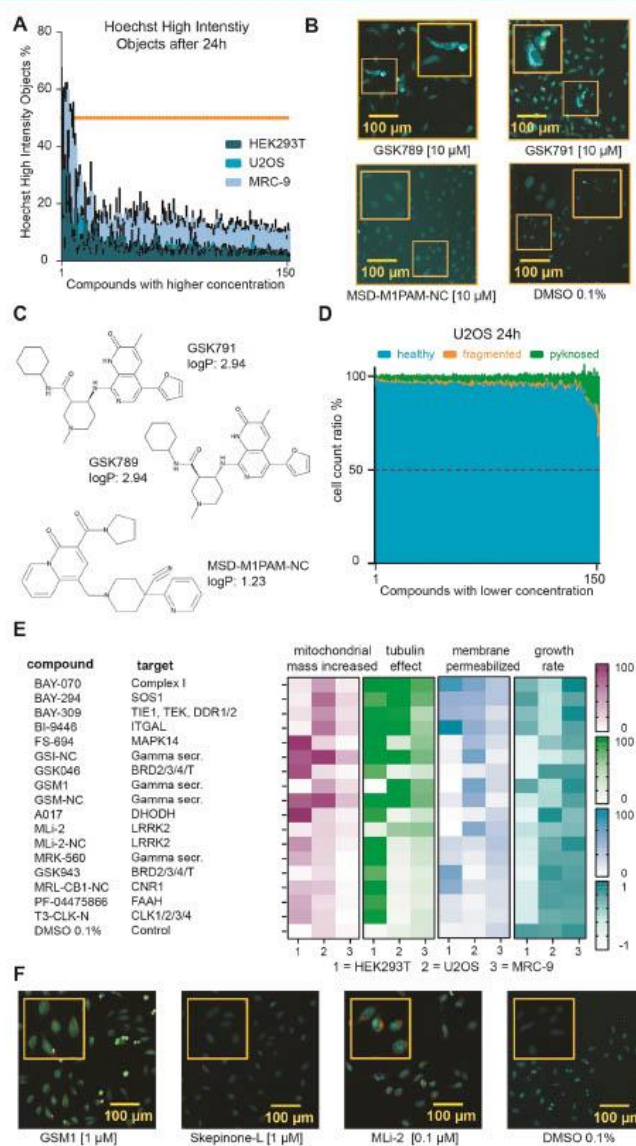
Figure 3. High-content multiplex assay in HEK293T, U2OS, and MRC-9 cells. All data are shown in Table S6. (A) Ratio of Hoechst High-Intensity Objects after screening at 10x higher concentrations for 24 h of higher concentration of compound exposure in HEK293T, U2OS, and MRC-9 cells. Error bars show SEM of biological duplicates. Property threshold at 50% marked orange. (B) Fluorescence image of stained (blue: DNA/nuclei, green: microtubule different, red: mitochondria content, magenta: Annexin V apoptosis marker) U2OS cells after screening at 10× higher concentration for 24 h of compound exposure (GSK791 10 μM, GSK789 10 μM) in comparison to U2OS cells treated with 0.1% DMSO. Compounds resulting in Hoechst High-Intensity Objects are highlighted. (C) Compound structures and corresponding log P values. (D) Cell

Figure 3. continued

count ratio of different nuclei gating after 24 h screening at recommended lower concentration of compound exposure in U2OS cells. Healthy nuclei count at 50% (viability threshold) is marked as a red line. (E) Heat map of phenotypic property ratios (increase in mitochondrial mass tubulin effect and membrane permeabilization) and the growth rate, which was calculated against the nontreated cell number, as described earlier,[31] after 24 h of HEK293T, U2OS, and MRC-9 cells exposed to 16 compounds that were marked as flagged after Multiplex analysis. Heat map shows the mean of two biological duplicates. (F) Fluorescence image of stained (blue: DNA/nuclei, green: microtubule different, red: mitochondria content, magenta: Annexin V apoptosis marker) U2OS cells after 24 h of compound exposure (GSM1 1 μM, Skepinone-L at 1 μM, and MLi-2 at 0.1 μM) in comparison to U2OS cells treated with 0.1% DMSO.

DCP set features two chemically unrelated DRD4 agonists, ABT-724 and UCSF924, including their respective controls.

Apart from on-target potency, target selectivity is the most critical parameter of a chemical probe and must be carefully assessed. Whenever possible, a family-specific selectivity assessment should be performed. All approved DCP probes have been profiled in comprehensive assay panels, including broad panels of pharmacologically active targets such as GPCRs, protein kinases, and ion channels (Tables S1–S5).

A screen[19] against 45 receptors (GPCRs and ion channels) was performed at 10 μM for all chemical probes and inactive controls at the National Institute of Mental Health (NIMH) Psychoactive Drug Screening Program (https://pdspdb.unc.edu/pdspWeb/), followed up with the determination of the K_i for primary results with greater than 50% inhibition. This data can be downloaded from the DCP website (Table S2).

Selectivity of DCP kinase probes was assessed in comprehensive kinase panels, such as KINOMEScan at DiscoverX, Eurofins,[20] containing 468 kinases, at the concentration recommended for cellular use. Hits from the screen were followed up either by determination of K_D/IC$_{50}$ values for primary values of less than 50% control inhibition and/or in an in-cell target engagement NanoBRET assay to determine the selectivity of the chemical probes in living cells (Tables S1 and S3).[21]

To illustrate the type of analysis and the extent of the data collected for each probe in the DCP set, we selected the chemical probe BAY-985[22] (Figure 2A), a potent inhibitor of the protein kinases TBK1 and IKBKE. This compound displayed 2.4 and 9.9% remaining activity against these two targets, respectively, when tested in a selectivity screen at 100 nM (Figure 2B). Additionally, MAP2K5 (MEK5) (8.5%), MAP3K19 (YSK4) (10%), SYK (31%), and STK17A (DRAK1) (40%) were identified as potential off-targets as well as binding to FLT3 (123 nM) was detected. A subsequent follow-up using the cellular target engagement assay Nano-BRET[23] revealed that BAY-985 engages in cells with TBK1 and IKBKE with 48 ± 5.1 and 420 ± 210 nM, respectively while binding to MAP2K5 and MAP3K19 was several orders of magnitude weaker (EC$_{50}$ > 50 μM) (Figure 2D). There is currently no NanoBRET assay available for SYK due to the lack of a suitable tracer, but a follow-up of the next off-targets revealed an EC$_{50}$ of 4.8 ± 2.0 μM for STK17A and 22 ± 3.9 μM for FLT3. The same targets were followed-up with the corresponding negative control BAY-440 (Figure 2A–C). BAY-440 showed a similar potency for FLT3 as for the probe molecule but lower potency for STK17A (Figure 2E). A more than 100-fold window to the target TBK1 and IKBKE compared to the chemical probe was confirmed.

Another insightful example for cellular off-target validation was the IRAK4 chemical probe GNE-2256, for which binding to FLT3, JAK1, JAK2, LRRK2, NTRK1, and NTRK2 has been reported in an *in vitro* kinase panel of 221 kinases (Table

S1).[24] Interestingly, in cells, these off-targets displayed more than 30-fold selectivity compared to the main target IRAK4 (Figure 2F), which had an EC$_{50}$ of 3 ± 0.3 nM in the NanoBRET assay. Additionally, both the probe and the inactive control compound showed a clean profile in a Differential-Scanning-Fluorimetry (DSF) assay comprising 86 kinases. Collectively, the DCP set includes selectivity profiling data obtained from target family assays, such as cellular NanoBRET data and DSF data. This data has been compiled in Figure S1 and Tables S4 and S5 and is available online.

General Cell Health Assessment and High-Content Multiplex Assay. Even a well-characterized chemical tool can produce misleading results when used at improper, *i.e.,* too high, concentrations due to the potential effects on proteins other than those included in selectivity profiling assays or other types of systemic effects. To minimize inappropriate use, we are distributing the DCP set to researchers in an assay ready format at 10,000-fold of the recommended concentration for cellular assays. Nevertheless, promiscuous interactions of a chemical probe with other proteins that have not been assessed cannot completely be excluded and may cause toxicity and unexpected phenotypic effects.[25] Therefore, we provide additional annotation of the probe compounds regarding their cellular quality. To assess the effects of the chemical probes on cell viability as well as on mitochondrial mass, tubulin morphology, and membrane permeability, a live-cell high-content screen was performed, as described previously.[26] All raw data and processed images have been deposited in Bioimage Archive (Accession number S-BIAD568#).

In total, 150 compounds were tested at the recommended (1×) as well as the 10-fold higher concentration (10×) in three different cell lines (U2OS, HEK293T, MRC-9 cells). To evaluate precipitation or auto-fluorescence of the compounds, Hoechst channel high-intensity objects were detected (Figure 3A). A 50% increased intensity was observed for seven DCP compounds at 10-fold of the recommended dose. Three of these compounds showed increased intensity in more than one cell line (Figure 3B). The BRD2/3/4/T chemical probe GSK789 and its negative control GSK791 both showed an increased intensity at 10-fold higher concentration, indicating a nuisance effect[27] of the compounds due to the structure. At the recommended concentration, this effect was not observed, supporting the advice to use the compounds not higher than at the recommended concentration in cellular assays. In addition, the negative control compound MSD-M1PAM-NC of the CHRM1 probe MSD-M1PAM showed 62% of Hoechst high-intensity objects in MRC-9 cells and 40% in U2OS cells, while no significant increase was detectable for the probe compound (Figure 3C). The logP values of MSD-M1PAM-NC (log P = 1.23) and MSD-M1PAM (log P = − 2.06) indicate that the solubility of the negative control compound might be limiting in cells at the higher concentration tested (Figure 3C).
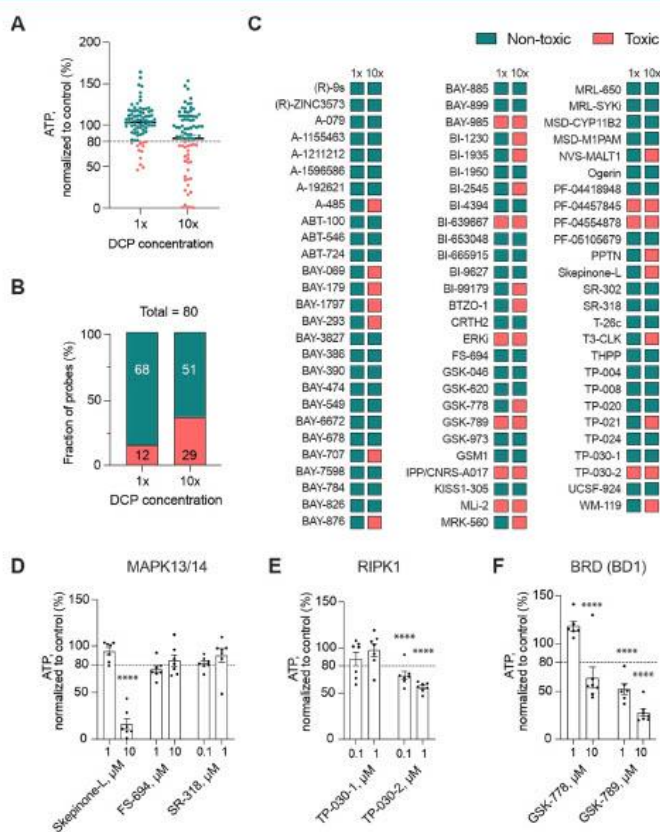
Figure 4. Hepatotoxicity screening of DCP compounds may assist to evaluate hits in other disease/phenotypic assays. (A) Overview of DCP hepatotoxic liability of DCP probes at 1× and 10× concentrations (see Methods). (B) Dose-dependent toxicity of the DCP compounds. (C) Graphical summary of hepatotoxicity of all 80 probes tested. Red boxes indicate that spheroid viability was reduced by >20% with statistical significance ($p < 0.05$, heteroscedastic two-tailed t-tests compared to vehicle controls). Green boxes indicate nontoxic compounds according to these criteria. Compounds for which toxicity was detected at low concentrations but not at higher concentrations were annotated as nontoxic. (D–F) Comparison of toxicity profiles between different DCP probes targeting the same protein, MAPK13/14 (D), RIPK1 (E), and BD1 (F). ****$p < 0.0001$ heteroscedastic two-tailed t-tests compared to vehicle controls.

The assessment of the cell viability based on the nuclear morphology over a time course of 24 h indicated that none of the tested compounds decreased cell viability to more than 50% at their recommended concentration in U2OS cells (Figure 3D) or nontransformed fibroblasts (MRC-9) (Table S6). In HEK293T cells, three compounds showed effects on viability after 24 h and decreased the count of healthy nuclei to up to 60% (GSI-NC, GSK046, NVS-MALT1) (Table S6). The BRD2/3/4/T probe GSK046 demonstrated an average of 42% healthy nuclei, and for the negative control of the γ secretase complex inhibitor MRK-560 GSI-NC, 45% healthy nuclei were detected after 24 h. At a 10-fold higher concentration, only treatment with BAY-309, the control compound of the Tie1,

TEK, DDR1/2 chemical probe BAY-826 resulted in less than 50% healthy cells after 24 h in all three cell lines. In contrast, BAY-826 had no effect at either 1× or 10× concentrations.

At the recommended concentration, only 17 of the 150 tested compounds showed one or more phenotypic effects after 24 h (Figure 3E). Twelve of the compounds affecting cell growth were negative controls and only five probe compounds were flagged in one or more of the cell lines (phenotypic effect higher than 50%). For example, treatment of cells with both γ secretase probes, MRK-560 and GSM1, resulted in an increase in mitochondrial mass (Figure 3E,F). Interestingly, both corresponding negative controls GSI-NC and GSM-NC induced the same phenotype (Figure 3E). The observed
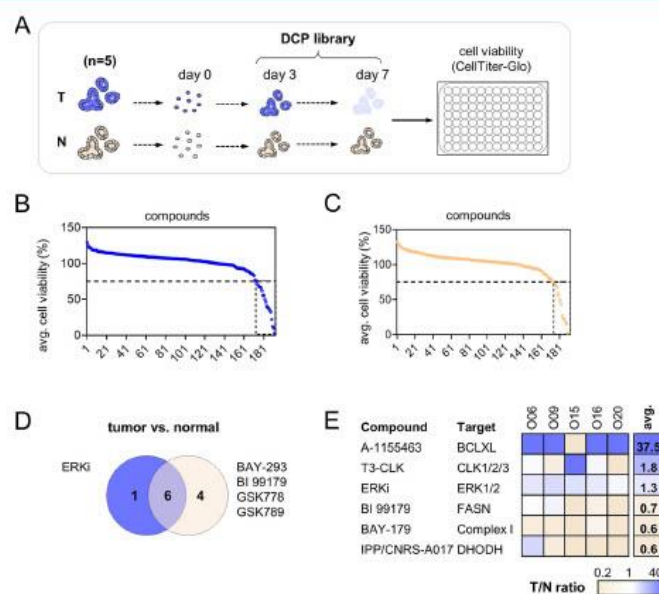
Figure 5. Chemical probe screen reveals tumor and normal specific hits. (A) Schematic overview of the compound screening workflow. (B/C) Average cell viability of all tumor (B) and normal lines (C). A 75% cell viability cut-off is applied for hit identification. (D) Venn diagram showing the overlap between top DCP hits in tumor organoids (ERKi) and normal organoids (BAY-293, BI99179, GSK778, GSK789). (E) Ratio of cell viability in the tumor vs normal is represented in the heat map, where the blue color indicates strong effects in tumor organoids and orange shows pronounced effects in normal organoid lines.

increase in mitochondrial mass could have been connected to the biology of γ secretase as it represents a key pharmacologic target of the Notch signaling pathway[28] and previous studies demonstrated that mitochondrial fission and Notch signaling are likely to be interdependent, which could explain the increase in mitochondrial mass of the chemical probes when binding to the γ secretase complex.[29] However, since the control compounds showed increased mitochondrial mass as well, this effect is likely not related to modulation of γ secretase activity, but might rather be caused by unspecific effects originating from the chemical scaffold, highlighting the importance of the use of accompanying negative control compounds. In addition, some of the compounds affected the tubulin read-out (Figure 3E). The other probes that were flagged include (a) the BRD2/3/4/T inhibitor GSK046, which influenced tubulin stains on average by 78% and increased mitochondrial mass to an average of 74% in HEK293T cells but showed no effect in the other cell lines (Figure 3E) and (b) MAPK14 inhibitor FS-694 that caused a change in tubulin stains (91%) and increased mitochondrial mass (84%) in HEK293T cells. MAPK14, also named p38α, is one of four p38 MAPKs and involved in the regulation of processes such as differentiation, autophagy, and apoptosis.[30] Interestingly, for the most selective p38 inhibitor Skepinone-L (Figure 3E,F), no effects were observed in our assay system, indicating that effects induced by FS-694 might be target-independent and rather due to chemical properties of the scaffold in HEK293T

cells. Finally, LRRK1 inhibitor MLi-2 (Figure 3E,F) and its control MLi-2-NC, both showed an increase in mitochondrial mass of more than 50% in one cell line. The unspecific effects, observed for some of the inactive control compounds, might be due to the lesser optimization and characterization of these compounds for the use in cells.

In summary, all probe compounds displayed a high cellular quality when used at their recommended concentrations in the tested cell lines. Nevertheless, these annotations should be considered basic guidance for the use of the set and are not a substitute to determine effects on cell health in the relevant system of the users.

Hepatotoxicity Screening in Primary Human Liver Spheroids. In addition to the general cell health assessment, we employed 3D cultures of primary human hepatocytes (PHH) that exhibit stable transcriptomic, proteomic, and metabolomic profiles in culture for multiple weeks.[32−34] This model has already been benchmarked extensively regarding the identification of hepatotoxic liabilities and has been shown to overall accurately distinguish toxic from nontoxic compounds.[33]

The hepatotoxicity of each DCP compound was evaluated at 1× and 10× recommended concentrations. Cell viability was assessed on day 7 after redosing at day 3 using ATP quantifications as a proxy. At the recommended concentration, only 12 out of 80 (15%) DCP compounds exhibited significant hepatotoxicity, while 29/80 were classified as toxic at 10-fold
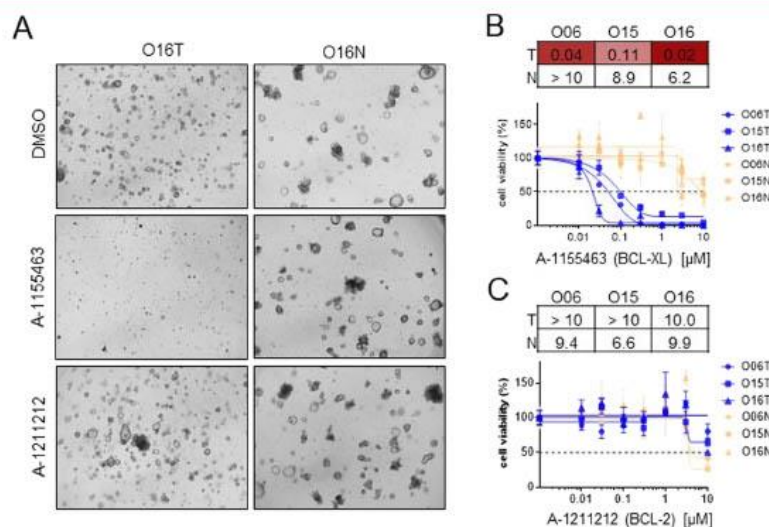
Figure 6. Validation of top hit compounds. (A–C) Tumor-selective sensitivity to BCL-XL inhibitor A-1155463. (A) Morphological overview of tumor and normal organoids (#O16) treated with 1 μM A-1155463 and A-1211212, resulting in pronounced toxicity of the BCL-XL inhibitor only in tumor cells. Organoids treated with DMSO or with the related BCL-2 inhibitor show unperturbed growth. Scale bars are 2 mm. (B/C) Dose-titration curves measured after 4 days of treatment with A-1155463 (B) and A-1211212 (C) are shown in three PDOs. IC$_{50}$ values are shown on top of each graph. Blue curves represent tumor organoids and orange normal organoids.

higher concentrations (Figure 4A–C and Table S7). Notably, we also find that some compounds increase spheroid ATP content. Potential causes could be alterations of cellular metabolism, induction of hepatocyte proliferation, or spheroid hypertrophy. However, further endpoints would be needed to evaluate the mechanism and potential relevance of these effects. In accordance with our phenotypic screening, these data demonstrate relatively low incidences of toxicity even after prolonged repeated dose exposure in primary human cells when probes are used at their recommended concentration.

We also noticed that certain probes exerted direct cellular toxicity over other probes targeting the same protein. For instance, Skepinone-L, FS694, and SR-318 are MAPK14 inhibitors with *in vitro* potency of similar magnitude, but contrary to the results in cell lines assessed by the multiplex assay, Skepinone-L exhibited clear hepatotoxicity at 10× concentrations, whereas FS694 and SR-318 were nontoxic (Figures 3F,4D and Table S7). These results confirm that the compound should be used within the recommended dose limit of up to 1 μM to avoid unspecific effects. Another interesting finding is that of the two analogues targeting RIPK1, TP-030-1 and TP-030-2; only the latter exhibited hepatotoxicity at all doses tested (Figure 4E). In contrast, for other targets, the different probes showed similar hepatotoxicity profiles, indicative of on-target toxicity, as exemplified by compounds targeting the bromodomain and extraterminal (BET) family BD1 (Figure 4F).

Screening of Colorectal Cancer-Specific Targets Using Patient-Derived Organoids. Finally, we evaluated the DCP library in patient-derived organoids (PDO) from colorectal cancer (CRC) tissue and matched normal-adjacent epithelium, as a model to identify shared and patient-specific tumor sensitivities. Tumor (T) and normal (N) organoids from five patients were tested using a semiautomatic platform. Organoids were seeded in a 96-well format as single cells and allowed to recover for 3 days, before drug exposure for 4 days and measurement of cell viability (Figure 5A). To normalize the seeding, colony formation assay was performed for each line (Figure S2A) and plate homogeneity was confirmed (Figure S2B). The DCP library was then screened at a single dose and in two experimental replicates for each line generating ~4000 data points (Table S8). In addition to negative control compounds, orthogonal chemical probes and inhibitors with a known mode of action were used in parallel to confirm the observed phenotype (Tables S8, S9 and Figure S3). Pearson correlation analysis showed robustness of experimental replicates (Figure S2C–E). Global comparison revealed the strongest correlation between each patient followed by the T/N status (Figure S3A). To score this, the average sensitivity was analyzed. At a 75% threshold, 7 and 10 hits were identified in T and N organoids, respectively, of which six DCP probes were shared between tumor and normal entities (Figures 5B–D and S3B,C). Most compounds demonstrated a prominent effect in all individuals; however, we also observed patient-specific sensitivities and resistances (Figure S3B,C). To study tumor selectivity, the T/N ratio was computed, which identified three tumor- and three normal-organoid selective chemical probes (Figure 5E). Notably, we observed a T/N ratio of nearly 40 for the BCL-XL inhibitor A-1155463.

We validated the top hits by dose-titration experiments. In three T/N pairs a pronounced antitumor toxicity was observed for the BCL-XL inhibitor A-1155463 with IC_{50}s between 0.02 and 0.11 μM that caused no toxicity in normal cell lines (Figure 6A,B). This is in line with our findings in the cellular health and hepatotoxicity screening, where A-1155463 showed no effect (Table S8). One tumor organoid pair (015T) was less sensitive to A-1155463 treatment, which was however also observed upon treatment with staurosporine, indicating a more generalized resistance phenotype (Table S9). Interestingly, inhibition of the related protein BCL-2 using A-1211212 neither affected tumor nor normal cells, indicating a highly specific vulnerability of CRC cells to BCL-XL (Figure 6C). These results reflect the antiapoptotic role of BCL-XL and data previously published,[35] which also showed a sensitivity of a primary colorectal spheroid culture to A-1155463 at an inhibitor concentration of about 100 nM. In contrast to BCL-XL, which is frequently overexpressed in CRC, BCL2 plays a secondary role in CRC[36] and an inverse expression pattern has been suggested for these two proteins.

In summary, our analysis revealed common tumor-specific vulnerabilities, indicating possibilities for personalized therapy targeting known and previously uncharacterized targets in CRC.

Here, we present a unique set of well-validated chemical probes. The physical compounds including inactive controls and characterization data are available via the associated website: https://www.sgc-ffm.uni-frankfurt.de/. In addition to recommendations regarding the appropriate concentrations provided with the set or individual probe and control compounds, we also provide additional annotations regarding the general toxicity of these chemical probes in model cell lines as well as in primary human hepatocyte organoids. We applied the compounds at the recommended concentration as well as a 10-fold higher concentration to assess the upper limit at which unspecific effects may occur to guide further use of the compounds. In most instances, no unexpected toxicities or cellular effects could be observed at the recommended dose, giving testimony to the high quality of probe compounds. Using a 10-fold higher than the recommended concentration resulted in some cases in toxic effects, which can be cell type specific as seen for the p38 inhibitors, highlighting the importance of appropriate dosing and underlining the need to evaluate toxicity of the compounds and the appropriate dosing for each cellular target. Testing the effect of the probe set on patient-derived organoids from colorectal cancer and matched normal-adjacent epithelium, we identified tumor-specific vulnerabilities. Using this screening method, we observed patient-specific tumor sensitivities and resistances, suggesting that a focused screening approach using the DCP library can be valuable to identify suitable targets for personalized medicine.

The DCP set is continuously expanding, fueled by donations from the pharmaceutical industry, biotech sector, and academia. Going forward, we are committed to maintaining the high standards of chemical probes added to the set by adhering to strict criteria for chemical probe validation, overseen by the internal and external review committees. Moreover, the open access nature of this resource should enable others to easily access high-quality tools and use them in their research without restrictions and large bureaucratic hurdles. We are grateful for the contributions from the participants and encourage scientists around the world to contribute high-quality paired probe compounds to this set, enriching it with binders for underexplored targets.

■ MATERIAL AND METHODS

DSF-Based Selectivity Screening against a Curated Kinase Library. The assay was performed as previously described.[37] Briefly, recombinant protein kinase domains at a concentration of 2 μM were mixed with 10 μM compound in a buffer containing 20 mM HEPES, pH 7.5, and 500 mM NaCl. SYPRO Orange (5000×, Invitrogen) was added as a fluorescent probe (1 μL per mL). Subsequently, temperature-dependent protein unfolding profiles were measured using the QuantStudio 5 realtime PCR machine (Thermo Fisher). Excitation and emission filters were set to 465 and 590 nm, respectively. The temperature was raised with a step rate of 3 °C/min. Data points were analyzed with the internal software (Thermal Shift Software Version 1.4, Thermo Fisher) using the Boltzmann equation to determine the inflection point of the transition curve.

NanoBRET Target Engagement Assays. The assay was performed as described previously.[21] In brief: Full-length kinase ORF (Promega) cloned in frame with a NanoLuc vector (as indicated in the table below) was transfected into HEK293T cells using FuGENE HD (Promega, E2312) and proteins were allowed to express for 20 h. Serially diluted inhibitor and NanoBRET Kinase Tracer (as indicated in the Table 2) were pipetted into white 384-well plates (Greiner 781207) using an ECHO 550 acoustic dispenser (Labcyte). The corresponding transfected cells were added and reseeded at a density of 2×10^5 cells/mL after trypsinization and resuspension in Opti-MEM without phenol red (Life Technologies). The system was allowed to equilibrate at 37 °C for 2 h and 5% CO_2 prior to BRET measurements. To measure BRET, NanoBRET NanoGlo Substrate + Extracellular NanoLuc Inhibitor (Promega, N2160) were added as per the manufacturer's protocol, and filtered luminescence was measured on a PHERAstar plate reader (BMG Labtech) equipped with a luminescence filter pair (450 nm BP filter (donor) and 610 nm LP filter (acceptor)). Competitive displacement data were normalized and then plotted using GraphPad Prism 9 software using a normalized three-parameter curve fit with the following equation

$$Y = 100/(1 + 10^{(X-\log IC50)})$$

Multiplex Assay. Viability assessment of three different cell lines was performed using a live-cell high-content screen as described previously by Tjaden et al.[26,38] In brief HEK293T (ATCC CRL-1573) and U2OS (ATCCHTB-96) were cultured in DMEM plus L-glutamine (high glucose) supplemented by 10% FBS (Gibco) and penicillin/streptomycin (Gibco). MRC-9 fibroblasts (ATCC CCL-2) were cultured in EMEM plus L-glutamine supplemented by 10% FBS (Gibco) and penicillin/streptomycin (Gibco). Cells were seeded at a density of 2000 cells per well in 384-well plates in a culture medium (cell culture microplate, PS, f-bottom, μClear, 781091, Greiner). Simultaneously, cells were stained with 60 nM Hoechst33342 (Thermo Scientific), 75 nM Mitotracker red (Invitrogen), 0.3 μL/well Annexin V Alexa Fluor 680 conjugate (Invitrogen), and 25 nL/well BioTracker 488 Green Microtubule Cytoskeleton Dye (EMD Millipore). Fluorescence and cellular shape were measured before compound treatment ((+)-JQ1), 12 and 24 h after compound exposure, respectively, using the CQ1 high-content confocal microscope (Yokogawa). The following setup parameters were used for image acquisition: Ex 405 nm/Em 447/60 nm, 500 ms, 50%; Ex 561 nm/Em 617/73 nm, 100 ms, 40%; Ex 488/Em 525/50 nm, 50 ms, 40%; bright field, 300 ms, 100% transmission, one centered field per well, 7 z stacks per well with 55 μm spacing. Images were analyzed using the CellPathfinder software (Yokogawa). Cells were detected as described previously[39] and gated using a machine learning algorithm. Data were normalized against the average of DMSO (0,1%) treated cells. Compounds were tested in duplicates and the complete screen was performed twice. SEM was calculated between biological duplicates.

Table 2

target	Nluc placement	target cat. no.	tracer	[tracer] (M)	tracer cat. no.
MAP3K5	C	kind gift of Promega	Tracer 7297	5.00×10^{-7}	kind gift of Promega
AKT1	C	NV2411	Tracer K10	1.00×10^{-6}	N2840
AKT2	C	NV1031	Tracer K10	5.00×10^{-7}	N2840
AKT3	C	kind gift of Promega	Tracer K10	1.00×10^{-6}	N2840
TEK	C	NV2151	Tracer K10	2.00×10^{-7}	N2840
TIE1	C	NV2171	Tracer K10	3.00×10^{-7}	N2840
STK10	N	NV4261	Tracer K10	5.00×10^{-7}	N2840
MAPK1	N	NV1641	Tracer K10	1.00×10^{-6}	N2840
TBK1	N	NV2131	Tracer K10	2.00×10^{-7}	N2840
SGK1	C	NV4221	Tracer K10	5.00×10^{-7}	N2840
PTK2	N	NV1921	Tracer K10	3.00×10^{-7}	N2840
PTK2B	C	NV1931	Tracer K10	3.00×10^{-7}	N2840
IKBKE	N	NV1431	Tracer K10	3.00×10^{-7}	N2840
STK17A	C	kind gift of Promega	Tracer K10	1.00×10^{-6}	N2840
FLT3	C	NV1391	Tracer K10	2.00×10^{-7}	N2840
MAP2K5	C	kind gift of Promega	Tracer K10	5.00×10^{-7}	N2840
MAP3K19	C	NV3461	Tracer K10	2.00×10^{-7}	N2840
IRAK4	C	NV1451	Tracer K10	2.00×10^{-7}	N2840
NTRK2	C	NV1821	Tracer K10	3.00×10^{-7}	N2840
JAK1	C	kind gift of Promega	Tracer K10	3.00×10^{-7}	N2840
JAK2	C	NV3291	Tracer K10	1.00×10^{-6}	N2840
DDR1	C	N2451	Tracer K4	3.20×10^{-8}	N2540
DDR2	C	NV1201	Tracer K4	1.00×10^{-7}	N2540
LYN	C	NV1551	Tracer K4	3.20×10^{-8}	N2540
EPHB6	C	kind gift of Promega	Tracer K4	1.00×10^{-6}	N2540
MAPK3	N	NV1671	Tracer K5	5.00×10^{-7}	N2530
NTRK1	C	NV1811	Tracer K5	2.00×10^{-7}	N2530
GSK3B	N	NV3201	Tracer K8	1.00×10^{-6}	N2820
GSK3A	N	NV3191	Tracer K8	2.00×10^{-7}	N2820
LRRK2	C	NV3401	Tracer K9	1.00×10^{-8}	N2830
CLK2	C	NV1141	Tracer K9	4.00×10^{-7}	N2830
CLK4	C	NV1151	Tracer K9	2.00×10^{-7}	N2830
MAP3K14	C	kind gift of Promega	Tracer K9	1.00×10^{-6}	N2830
TTK	C	NV2191	Tracer K9	7.00×10^{-7}	N2830

Organoid Culture. Resection samples from colorectal cancer patients were provided by the University Cancer Center Frankfurt (UCT). All materials were collected after prior written informed consent as part of the interdisciplinary Biobank and Database Frankfurt (iBDF) and the study was approved by the institutional review board of the University Hospital Frankfurt (ethics vote: 4/09; project numbers SGI-12-2018 and SGI-10-2022). Patient-derived organoids were established and cultured as previously described.[40] Briefly, medium containing advanced DMEM/F12 supplemented with 10 mM HEPES, 1× Glutamax, 1× penicillin/streptomycin, 2% B27, 12.5 mM N-acetylcysteine, 500 nM A83-01 (R&D Systems), 10 μM SB202190 (Sigma-Aldrich), 20% R-spondin 1 conditioned medium, 10% Noggin conditioned medium, 50 ng/mL human EGF (Peprotech), and Wnt surrogate (35 ng/mL, # N001-0.5 mg, ImmunoPrecise) was used for both tumor and normal organoids. Y-27632 (10 μM) was added to the medium for the first 3 days after passaging.

Compound Screen in Patient-Derived Organoids. To standardize the input organoid number for compound screen, we performed colony formation assay for each tumor-normal pair by seeding 500, 1000, 2000, and 5000 live cells/well in 15 μL 90% Matrigel (Corning) in 96-well round bottom plates (Sarstedt). A 100 μL/well medium was added to the cells and the plate was sealed with Breath-easy membranes (Sigma-Aldrich) to avoid medium evaporation. Organoids were beforehand dissociated enzymatically with Accutase (Thermo Fischer), and single cells were filtered (40 μm, 542040, Greiner). Morphological images from three wells were collected (2× objective, EVOS Fl, Life Technologies) after 7 days and

colonies were counted with ImageJ software. Linear regression analysis was performed with GraphPad Prism to estimate the number of organoids/2000 input cells. To evaluate the seeding uniformity, the coefficient of variation was calculated on the raw data from the entire 96-well plate. Pearson correlation was used to calculate the reproducibility between two replicate plates from normalized values of each replicate.

The DCP library was stored at −80 °C in 5 μL volume in DMSO in a 96-well plate format at the following concentrations: 1, 2, or 10 mM. On the day of treatment, the stock plates were further diluted to a final concentration of 0.1, 0.2, or 1 μM. Tumor and normal organoids were harvested, singularized by enzymatic digestion, filtered, and counted. The seeding organoid cell number was calculated for each line, to optimally obtain 300 org/well. Cells were embedded in 15 μL of Matrigel and seeded in 96-well round bottom plates, as described above. The medium was changed after 3 days and the library was administered with an automatic multichannel pipette (Integra Mini 96). The experiment was stopped after 4 days of treatment, representative morphological images were taken, and cell viability was assessed using CellTiter-Glo assay (Promega) following the manufacturer's recommendations. Luminescence was measured on a SpectraMax iD3 Microplate Reader (Molecular Devices). The data were normalized to the DMSO control and centered for each replicate plate. The plate centering factor is indicated in the raw data in Table S8.

For validation of the hit compounds, three matched tumor and normal organoids were used. DMSO stock compounds (10 mM) were stored at −20 °C. Organoid cells were seeded as above, in

technical triplicate, in 96-well plates. On day 3, the medium was changed and compounds were dispensed with D300e digital dispenser (Tecan) in 7-point dilutions ranging from 0.01 to 10 μM in a logarithmic scale. The DMSO content was normalized to the highest class volume in all wells, not exceeding 1% of the final volume. Cell viability was assessed after 4 days using the CellTiter-Glo (as above). Raw data were normalized to the average of 12 DMSO controls present per plate. Dose–response analysis and IC_{50} data were generated in GraphPad Prism.

3D Liver Spheroid Culture. PHH were seeded in ultralow attachment plates (faCellitate) with densities of 1500 cells per well in William's E medium (Gibco) supplemented with 2 mM L-glutamine, 100U/mL penicillin, 100 μg/mL streptomycin, 10 μg/mL insulin, 5.5 μg/mL transferrin, 6.7 ng/mL sodium selenite, 100 nM dexamethasone, and 10% fetal bovine serum (FBS) as previously reported.[32] Self-aggregation occurred within 1 week. Subsequently, spheroids were treated for 7 days with FBS-free media containing the DCP compounds at two concentrations ($1\times$ and $10\times$). Cells were redosed at day 3 and cell viability was measured at day 7. Cell viability was assessed by measuring cellular ATP using the CellTiter-Glo Luminescent Cell Viability Assay kit (Promega) with luminescence reading (RLU). Toxicity was defined on the basis of heteroscedastic two-tailed t-tests between treated spheroids and controls and compounds were classified as hepatotoxic if they resulted in a statistically significant reduction of viability below 80% of control levels.[33]

■ ASSOCIATED CONTENT

Data Availability Statement

All data are readily accessible via the accompanying website and database https://www.sgc-ffm.uni-frankfurt.de/.

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acschembio.2c00877.

> Representative curves for NanoBRET on-target and off-target profiling of donated chemical probes and negative controls (Figure S1); assay development for DCP screening in PDO models (Figure S2); and compound screening in patient-derived organoids (Figure S3) (PDF)
>
> Annotated list of 97 donated chemical probes including the target, affinity, and selectivity data, as well as recommendations on usage and control compounds (Table S1) (XLSX)
>
> Receptor screen (GPCRs and ion channels) for donated probes and negative controls (Table S2) (XLSX)
>
> Kinome-wide screening of donated probes and negative controls (Table S3) (XLSX)
>
> NanoBRET on-target and off-target profiling of donated chemical probes and negative controls (Table S4) (XLSX)
>
> DSF kinase panel comprising 86 kinases performed for the IRAK4 chemical probe GNE 225 and its inactive control GNE-6689 (Table S5) (XLSX)
>
> Cell health screen assessed by multiplex assay for donated probes and negative controls (Table S6) (XLSX)
>
> Toxicity effects of compounds in HEK293T, U2OS, and MRC-9 cells vs primary human hepatocytes (Table S7) (XLSX)
>
> Donated chemical probes screening in patient-derived organoids (Table S8) (XLSX)
>
> Summarized data of the donated chemical probes screening in patient-derived organoids (Table S9) (XLSX)

■ AUTHOR INFORMATION

Corresponding Authors

Susanne Müller – Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Buchmann Institute for Molecular Life Sciences and Structural Genomics Consortium (SGC), 60438 Frankfurt am Main, Germany; ◉ orcid.org/0000-0003-2402-4157; Email: tredup@pharmchem.uni-frankfurt.de

Claudia Tredup – Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Buchmann Institute for Molecular Life Sciences and Structural Genomics Consortium (SGC), 60438 Frankfurt am Main, Germany; Email: susanne.mueller-knapp@bmls.de

Authors

Benardina Ndreshkjana – Georg-Speyer-Haus, Institute for Tumor Biology and Experimental Therapy, 60596 Frankfurt am Main, Germany

Natalie S. Schneider – Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Buchmann Institute for Molecular Life Sciences and Structural Genomics Consortium (SGC), 60438 Frankfurt am Main, Germany

Amelie Tjaden – Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Buchmann Institute for Molecular Life Sciences and Structural Genomics Consortium (SGC), 60438 Frankfurt am Main, Germany

Aurino M. Kemas – Department of Physiology and Pharmacology, Karolinska Institutet, 171 65 Stockholm, Sweden

Sonia Youhanna – Department of Physiology and Pharmacology, Karolinska Institutet, 171 65 Stockholm, Sweden

Volker M. Lauschke – Department of Physiology and Pharmacology, Karolinska Institutet, 171 65 Stockholm, Sweden; Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology, 70376 Stuttgart, Germany; University of Tübingen, 72074 Tübingen, Germany; ◉ orcid.org/0000-0002-1140-6204

Benedict-Tilman Berger – Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Buchmann Institute for Molecular Life Sciences and Structural Genomics Consortium (SGC), 60438 Frankfurt am Main, Germany; ◉ orcid.org/0000-0002-3314-2617

Andreas Krämer – Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Buchmann Institute for Molecular Life Sciences and Structural Genomics Consortium (SGC), 60438 Frankfurt am Main, Germany; Frankfurt Cancer Institute, Goethe University, Frankfurt am Main 60596, Germany

Lena M. Berger – Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Buchmann Institute for Molecular Life Sciences and Structural Genomics Consortium (SGC), 60438 Frankfurt am Main, Germany; ◉ orcid.org/0000-0002-7835-8067

Sandra Röhm – Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Buchmann Institute for Molecular Life Sciences and Structural Genomics Consortium (SGC), 60438

Frankfurt am Main, Germany; ◉ orcid.org/0000-0003-3999-712X

Stefan Knapp − *Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Buchmann Institute for Molecular Life Sciences and Structural Genomics Consortium (SGC), 60438 Frankfurt am Main, Germany; Frankfurt Cancer Institute, Goethe University, Frankfurt am Main 60596, Germany*

Henner F. Farin − *Georg-Speyer-Haus, Institute for Tumor Biology and Experimental Therapy, 60596 Frankfurt am Main, Germany; Frankfurt Cancer Institute, Goethe University, Frankfurt am Main 60596, Germany; German Cancer Consortium (DKTK), Heidelberg 69120, Germany; German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acschembio.2c00877

Author Contributions

¶C.T. and B.N. contributed equally to this work.

Notes

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Fedorov, O.; Muller, S.; Knapp, S. The (un)targeted cancer kinome. *Nat Chem Biol* 2010, 6, 166−169.

(2) Arrowsmith, C. H.; Audia, J. E.; Austin, C.; Baell, J.; Bennett, J.; Blagg, J.; Bountra, C.; Brennan, P. E.; Brown, P. J.; Bunnage, M. E.; et al. The promise and peril of chemical probes. *Nat Chem Biol* 2015, 11, 536−541.

(3) Blagg, J.; Workman, P. Choose and Use Your Chemical Probe Wisely to Explore Cancer Biology. *Cancer Cell* 2017, 32, 9−25.

(4) Edwards, A. M.; Bountra, C.; Kerr, D. J.; Willson, T. M. Open access chemical and clinical probes to support drug discovery. *Nat Chem Biol* 2009, 5, 436−440.

(5) Bunnage, M. E.; Chekler, E. L.; Jones, L. H. Target validation using chemical probes. *Nat Chem Biol* 2013, 9, 195−199.

(6) Baell, J. B.; Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* 2010, 53, 2719−2740.

(7) Békés, M.; Langley, D. R.; Crews, C. M. PROTAC targeted protein degraders: the past is prologue. *Nat. Rev. Drug Discovery* 2022, 21, 181−200.

(8) Antolin, A. A.; Workman, P.; Al-Lazikani, B. Public resources for chemical probes: the journey so far and the road ahead. *Future Med. Chem.* 2021, 13, 731−747.

(9) Antolin, A. A.; Tym, J. E.; Komianou, A.; Collins, I.; Workman, P.; Al-Lazikani, B. Objective, Quantitative, Data-Driven Assessment of Chemical Probes. *Cell Chem. Biol.* 2018, 25, 194−205. e195.

(10) Antolin, A. A.; Sanfelice, D.; Crisp, A.; Villasclaras Fernandez, E.; Mica, I. L.; Chen, Y.; Collins, I.; Edwards, A.; Muller, S.; Al-Lazikani, B.; et al. The Chemical Probes Portal: an expert review-based public resource to empower chemical probe assessment, selection and use. *Nucleic Acids Res.* 2022, 51, D1492−D1502.

(11) Mullard, A. Boehringer Ingelheim experiments with open-access chemical probes. *Nat. Rev. Drug Discovery* 2018, 17, 7.

(12) Gollner, A.; Koster, M.; Nicklin, P.; Trieselmann, T.; Klein, E.; Vlach, J.; Heine, C.; Grundl, M.; Ramharter, J.; Wyatt, D.; et al. opnMe.com: a digital initiative for sharing tools with the biomedical research community. *Nat Rev. Drug Discovery* 2022, 21, 475−476.

(13) Wells, C. I.; Al-Ali, H.; Andrews, D. M.; Asquith, C. R. M.; Axtman, A. D.; Dikic, I.; Ebner, D.; Ettmayer, P.; Fischer, C.; Frederiksen, M.; et al. The Kinase Chemogenomic Set (KCGS): An Open Science Resource for Kinase Vulnerability Identification. *Int. J. Mol. Sci.* 2021, 22, No. 566.

(14) Müller, S.; Ackloo, S.; Arrowsmith, C. H.; Bauser, M.; Baryza, J. L.; Blagg, J.; Bottcher, J.; Bountra, C.; Brown, P. J.; Bunnage, M. E.; et al. Donated chemical probes for open science. *eLife* 2018, 7, No. e34311.

(15) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J. W.; da Silva Santos, L. B.; Bourne, P. E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 2016, 3, No. 160018.

(16) Proj, M.; Knez, D.; Sosic, I.; Gobec, S. Redox active or thiol reactive? Optimization of rapid screens to identify less evident nuisance compounds. *Drug Discovery Today* 2022, 27, 1733−1742.

(17) Baell, J.; Walters, M. A. Chemistry: Chemical con artists foil drug discovery. *Nature* 2014, 513, 481−483.

(18) Lee, J.; Schapira, M. The Promise and Peril of Chemical Probe Negative Controls. *ACS Chem Biol* 2021, 16, 579−585.

(19) Besnard, J.; Ruda, G. F.; Setola, V.; Abecassis, K.; Rodriguiz, R. M.; Huang, X. P.; Norval, S.; Sassano, M. F.; Shin, A. I.; Webster, L. A.; et al. Automated design of ligands to polypharmacological profiles. *Nature* 2012, 492, 215−220.

(20) Davis, M. I.; Hunt, J. P.; Herrgard, S.; Ciceri, P.; Wodicka, L. M.; Pallares, G.; Hocker, M.; Treiber, D. K.; Zarrinkar, P. P. Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* 2011, 29, 1046−1051.

(21) Vasta, J. D.; Corona, C. R.; Wilkinson, J.; Zimprich, C. A.; Hartnett, J. R.; Ingold, M. R.; Zimmerman, K.; Machleidt, T.; Kirkland, T. A.; Huwiler, K. G.; et al. Quantitative, Wide-Spectrum Kinase Profiling in Live Cells for Assessing the Effect of Cellular ATP on Target Engagement. *Cell Chem. Biol.* 2018, 25, 206−214. e211.

(22) Lefranc, J.; Schulze, V. K.; Hillig, R. C.; Briem, H.; Prinz, F.; Mengel, A.; Heinrich, T.; Balint, J.; Rengachari, S.; Irlbacher, H.; et al. Discovery of BAY-985, a Highly Selective TBK1/IKKepsilon Inhibitor. *J. Med. Chem.* 2020, 63, 601–612.

(23) Robers, M. B.; Friedman-Ohana, R.; Huber, K. V. M.; Kilpatrick, L.; Vasta, J. D.; Berger, B. T.; Chaudhry, C.; Hill, S.; Muller, S.; Knapp, S.; Wood, K. Quantifying Target Occupancy of Small Molecules Within Living Cells. *Annu. Rev. Biochem.* 2020, 89, 557–581.

(24) Rajapaksa, N. S.; Gobbi, A.; Drobnick, J.; Do, S.; Kolesnikov, A.; Liang, J.; Chen, Y.; Sujatha-Bhaskar, S.; Huang, Z.; Brightbill, H.; et al. Discovery of Potent Benzolactam IRAK4 Inhibitors with Robust in Vivo Activity. *ACS Med. Chem. Lett.* 2020, 11, 327–333.

(25) Price, D. A.; Blagg, J.; Jones, L.; Greene, N.; Wager, T. Physicochemical drug properties associated with in vivo toxicological outcomes: a review. *Expert Opin. Drug Metab. Toxicol.* 2009, 5, 921–931.

(26) Tjaden, A.; Chaikuad, A.; Kowarz, E.; Marschalek, R.; Knapp, S.; Schroder, M.; Muller, S. Image-Based Annotation of Chemogenomic Libraries for Phenotypic Screening. *Molecules* 2022, 27, No. 1439.

(27) Chakravorty, S. J.; Chan, J.; Greenwood, M. N.; Popa-Burke, I.; Remlinger, K. S.; Pickett, S. D.; Green, D. V. S.; Fillmore, M. C.; Dean, T. W.; Luengo, J. I.; Macarrón, R. Nuisance Compounds, PAINS Filters, and Dark Chemical Matter in the GSK HTS Collection. *SLAS Discovery* 2018, 23, 532–545.

(28) Aster, J. C.; Pear, W. S.; Blacklow, S. C. The Varied Roles of Notch in Cancer. *Annu. Rev. Pathol.* 2017, 12, 245–275.

(29) Chen, L.; Zhang, J.; Lyu, Z.; Chen, Y.; Ji, X.; Cao, H.; Jin, M.; Zhu, J.; Yang, J.; Ling, R.; et al. Positive feedback loop between mitochondrial fission and Notch signaling promotes survivin-mediated survival of TNBC cells. *Cell Death Dis.* 2018, 9, 1050.

(30) Desideri, E.; Vegliante, R.; Cardaci, S.; Nepravishta, R.; Paci, M.; Ciriolo, M. R. MAPK14/p38alpha-dependent modulation of glucose metabolism affects ROS levels and autophagy during starvation. *Autophagy* 2014, 10, 1652–1665.

(31) Hafner, M.; Niepel, M.; Chung, M.; Sorger, P. K. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nat. Methods* 2016, 13, 521–527.

(32) Bell, C. C.; Hendriks, D. F.; Moro, S. M.; Ellis, E.; Walsh, J.; Renblom, A.; Fredriksson Puigvert, L.; Dankers, A. C.; Jacobs, F.; Snoeys, J.; et al. Characterization of primary human hepatocyte spheroids as a model system for drug-induced liver injury, liver function and disease. *Sci. Rep.* 2016, 6, No. 25187.

(33) Vorrink, S. U.; Zhou, Y.; Ingelman-Sundberg, M.; Lauschke, V. M. Prediction of Drug-Induced Hepatotoxicity Using Long-Term Stable Primary Hepatic 3D Spheroid Cultures in Chemically Defined Conditions. *Toxicol. Sci.* 2018, 163, 655–665.

(34) Bell, C. C.; Lauschke, V. M.; Vorrink, S. U.; Palmgren, H.; Duffin, R.; Andersson, T. B.; Ingelman-Sundberg, M. Transcriptional, Functional, and Mechanistic Comparisons of Stem Cell-Derived Hepatocytes, HepaRG Cells, and Three-Dimensional Human Hepatocyte Spheroids as Predictive In Vitro Systems for Drug-Induced Liver Injury. *Drug Metab. Dispos.* 2017, 45, 419–429.

(35) Ramesh, P.; Di Franco, S.; Atencia Taboada, L.; Zhang, L.; Nicotra, A.; Stassi, G.; Medema, J. P. BCL-XL inhibition induces an FGFR4-mediated rescue response in colorectal cancer. *Cell Rep.* 2022, 38, No. 110374.

(36) Ramesh, P.; Medema, J. P. BCL-2 family deregulation in colorectal cancer: potential for BH3 mimetics in therapy. *Apoptosis* 2020, 25, 305–320.

(37) (a) Fedorov, O.; Niesen, F. H.; Knapp, S. Kinase inhibitor selectivity profiling using differential scanning fluorimetry. *Methods Mol. Biol.* 2012, 795, 109–118. (b) Krämer, A.; Kurz, C. G.; Berger, B. T.; Celik, I. E.; Tjaden, A.; Greco, F. A.; Knapp, S.; Hanke, T. Optimization of pyrazolo[1,5-a]pyrimidines lead to the identification of a highly selective casein kinase 2 inhibitor. *Eur. J. Med. Chem.* 2020, 208, 112770.

(38) Tjaden, A.; Giessmann, R. T.; Knapp, S.; Schroder, M.; Muller, S. High-content live-cell multiplex screen for chemogenomic compound annotation based on nuclear morphology. *STAR Protoc.* 2022, 3, No. 101791.

(39) Howarth, A.; Schroder, M.; Montenegro, R. C.; Drewry, D. H.; Sailem, H.; Millar, V.; Muller, S.; Ebner, D. V. HighVia-A Flexible Live-Cell High-Content Screening Pipeline to Assess Cellular Toxicity. *SLAS Discovery* 2020, 25, 801–811.

(40) van de Wetering, M.; Francies, H. E.; Francis, J. M.; Bounova, G.; Iorio, F.; Pronk, A.; van Houdt, W.; van Gorp, J.; Taylor-Weiner, A.; Kester, L.; et al. Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell* 2015, 161, 933–945.

Appendix F: Illuminating the Dark: Highly Selective Inhibition of Serine/Threonine Kinase 17a with Pyrazolo[1,5-a]Pyrimidine-Based Macrocycles

Contribution Menge, A.: Viability assessment using Multiplex high-content live-cell assay, data evaluation, figure preparation, manuscript editing.

Article

Illuminating the Dark: Highly Selective Inhibition of Serine/Threonine Kinase 17A with Pyrazolo[1,5-*a*]pyrimidine-Based Macrocycles

Christian G. Kurz, Franziska Preuss, Amelie Tjaden, Martin Cusack, Jennifer Alisa Amrhein, Deep Chatterjee, Sebastian Mathea, Lena Marie Berger, Benedict-Tilman Berger, Andreas Krämer, Michael Weller, Tobias Weiss, Susanne Müller, Stefan Knapp,* and Thomas Hanke*

Cite This: *J. Med. Chem.* 2022, 65, 7799−7817

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Serine/threonine kinase 17A (death-associated protein kinase-related apoptosis-inducing protein kinase 1—DRAK1) is a part of the death-associated protein kinase (DAPK) family and belongs to the so-called dark kinome. Thus, the current state of knowledge of the cellular function of DRAK1 and its involvement in pathophysiological processes is very limited. Recently, DRAK1 has been implicated in tumorigenesis of glioblastoma multiforme (GBM) and other cancers, but no selective inhibitors of DRAK1 are available yet. To this end, we optimized a pyrazolo[1,5-*a*]pyrimidine-based macrocyclic scaffold. Structure-guided optimization of this macrocyclic scaffold led to the development of CK156 (34), which displayed high in vitro potency (K_D = 21 nM) and selectivity in kinomewide screens. Crystal structures demonstrated that CK156 (34) acts as a type I inhibitor. However, contrary to studies using genetic knockdown of DRAK1, we have seen the inhibition of cell growth of glioma cells in 2D and 3D culture only at low micromolar concentrations.

Illuminating the Dark Kinome

CK156 a Chemical Tool for DRAK1:

Cellular potency
IC_{50} on DRAK1 = 182 nM
Selectivity Score (35) = 0.005
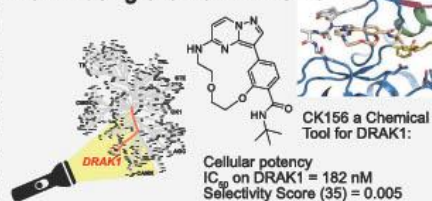
1. INTRODUCTION

Since the first clinical approval of the small-molecule kinase inhibitor imatinib in 2001, the number of kinase inhibitors entering clinical trials has increased dramatically. Kinases are among the most promising drug target families for various types of human diseases because of their critical function in cellular signaling pathways.[1−5] Besides their role in oncology, kinase inhibitors have now also emerged as therapeutically approved agents for the treatment of other diseases such as autoimmune and inflammatory diseases.[6] All currently approved inhibitors target the enzymatic activity of the kinase, thus preventing phosphorylation of downstream signaling events by orthosteric or allosteric inhibition.[7] The kinase catalytic process is dependent on ATP, which binds to a highly conserved active site located between the N- and C-lobe of the catalytic domain. Consequently, the development of highly selective ATP-mimetic inhibitors is challenging.[1,8] Typical ATP-mimetic kinase inhibitors are often flat hydrophobic molecules with many rotatable bonds. One possibility to improve and control shape complementarity and to lock a bioactive inhibitor conformation is the introduction of macrocyclization. Macrocyclic kinase inhibitors typically contain an ATP-mimetic pharmacophore, which is constrained by a linker of 12 or more atoms that connects flexible moieties

in acyclic conventional inhibitors.[9,10] As a result of the conformational restriction, the entropic costs during binding can be reduced, which presumably leads to an increased binding affinity to the target protein.[11] Despite the loss of flexibility, macrocycles are far from being rigid and can therefore also interact with dynamic binding sites in proteins.[12] Moreover, macrocyclization strategies not only allow the modulation of inhibitor potency and selectivity, but they also enable optimization of physiochemical and pharmacokinetic properties such as solubility, lipophilicity, metabolic stability, and bioactivity.[13]

Serine/threonine kinase 17A (STK17A), also known as death-associated protein kinase-related apoptosis-inducing protein kinase 1 (DRAK1),[14] belongs to the family of death-associated protein kinases (DAPKs) along with DAPK1, DAPK2, DAPK3, and DRAK2. DAPK1 is the most studied kinase of this small family. DAPK1 regulates many different
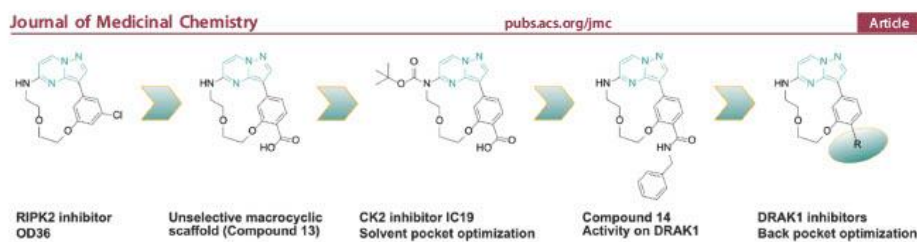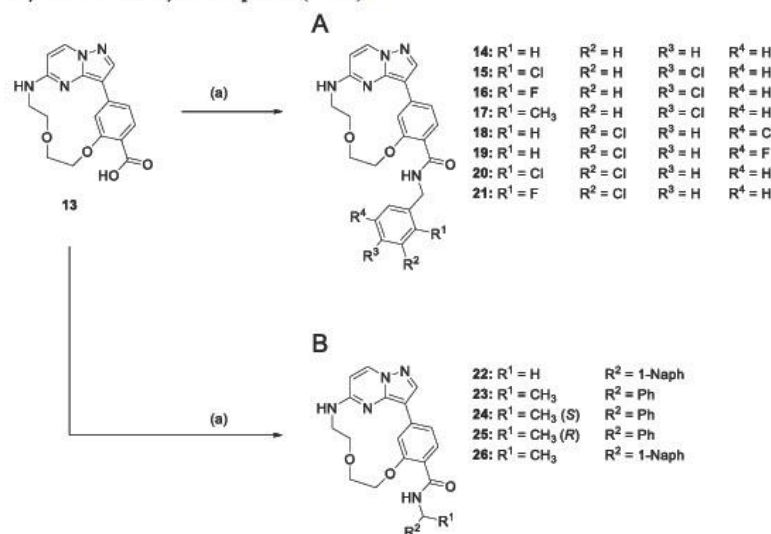
Figure 1. Schematic overview of the development of DRAK1 inhibitors by exploring the back pocket. Pyrazolo[1,5-a]pyrimidine-based RIPK2 inhibitor OD36 was used as a starting point to develop an unselective macrocyclic core scaffold. Optimization of interactions in the solvent pocket resulted in the highly selective CK2 inhibitor IC19.[31] Modification of the back pocket interaction in this study identified the lead structure (14), which lead to selective DRAK1 inhibitors.
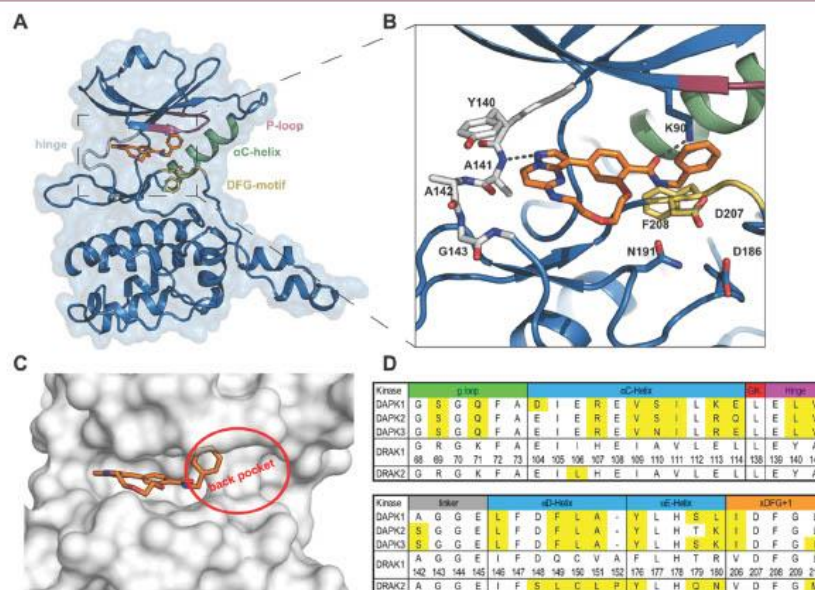
Figure 2. Structural features of the DRAK1 complex with 14 (A) overall view of a monomer of the co-crystal structure of DRAK1 with the macrocyclic lead structure 14 (PDB: 7QUE). The main structural elements are labeled and 14 is shown in stick representation (orange carbon atoms). The hinge region is highlighted in gray, the P-loop in magenta, the αC-helix in green, and the DFG motif in yellow. (B) Binding mode of 14 using the same color-coding for structural elements as described in (A). The ligand binds the hinge forming a hydrogen bond (dashed line) with the backbone nitrogen of A141. An additional hydrogen bond interaction was formed between the carbonyl oxygen of the amide and the side chain amine of the catalytic lysine (K90). The benzylamide motif extends into the back pocket (C). (D) Structural alignment of key residues in the ATP binding pocket of the DAPK family. Residues highlighted in yellow marked differences compared to DRAK1.

cellular processes, such as autophagy[15] or apoptosis,[16] and additional functions have been reported.[17] The kinase domains of DAPK2 and DAPK3 are highly conserved compared to DAPK1 sharing 83 and 80% identity at the amino acid level, respectively, whereas the kinase domains of DRAK1 and DRAK2 are only 48 and 51% identical in sequence. The closest family member of DRAK1 is DRAK2 with 66% sequence identity in the kinase domain, whereas the other DAPKs (1–3) share only 46%, respectively, 45 and 45% sequence identity in the kinase domain to DRAK1 (sequence alignment was performed using Clusta Omega software[18]). However, while the N-terminal catalytic kinase domains share high similarity, the C-terminal regulatory domains of DAPKs differ substantially,[19] responding to different stimuli for induced cell death in a context-dependent way.[20] For example, in testicular cancer cells, cell death is induced by DRAK1 as a result of reactive oxygen species, induced by cisplatin in a p53-dependent manner.[21] DRAK1 is mainly localized in the nucleus where it interacts with p53. Nevertheless, DRAK1 can translocate into the cytosol upon the activation of protein kinase C (PKC) and

Scheme 1. Synthesis of Macrocyclic Compounds (14−26)[a]

A

14: R^1 = H	R^2 = H	R^3 = H	R^4 = H
15: R^1 = Cl	R^2 = H	R^3 = Cl	R^4 = H
16: R^1 = F	R^2 = H	R^3 = Cl	R^4 = H
17: R^1 = CH_3	R^2 = H	R^3 = Cl	R^4 = H
18: R^1 = H	R^2 = Cl	R^3 = H	R^4 = Cl
19: R^1 = H	R^2 = Cl	R^3 = H	R^4 = F
20: R^1 = Cl	R^2 = Cl	R^3 = H	R^4 = H
21: R^1 = F	R^2 = Cl	R^3 = H	R^4 = H

B

22: R^1 = H	R^2 = 1-Naph
23: R^1 = CH_3	R^2 = Ph
24: R^1 = CH_3 (S)	R^2 = Ph
25: R^1 = CH_3 (R)	R^2 = Ph
26: R^1 = CH_3	R^2 = 1-Naph

[a](A) Benzyl amide-based derivatives and (B) phenylethan- and naphtylmethyl/ethylamide-based derivatives. Reagents and general conditions: (a) (1) HATU, DIPEA, dry DMF, argon atmosphere, 1 h, rt. (2) Corresponding amine, 18 h, rt.

DRAK1 can also be found in mitochondria interacting with the anti-apoptotic oncoprotein adenine nucleotide translocator 2 (ANT2).[22] In NIH3T3 cells, an overexpression of DRAK1 induces morphological changes that are characteristic for apoptosis.[20] Overexpression of cytoplasmic DRAK1 in head and neck squamous cell carcinoma (HNSCC) leads to the negative regulation of a tumor suppressor function of TGF-β1. This is achieved by binding to Smad3, which is essential for transforming growth factor beta (TGF-β1) signaling.[23] Compared to the normal brain tissue, DRAK1 is overexpressed in multiple gliomas, with the highest expression levels occurring in glioblastoma multiforme (GBM), promoting cell proliferation, migration, and resistance to genotoxic agents in GBM.[24,25]

To date, no selective DRAK1 inhibitors have been reported, but DRAK1 has been identified as an off-target for several inhibitors directed at other kinases. For instance, the highly potent MNK1/2 inhibitor tomivosertib (eFT508) (IC_{50}: 2.4 and 1 nM, respectively, on MNK1/2) inhibits also DRAK1 with an IC_{50} of 131 nM.[26] Other examples are the aurora A/B inhibitor AT9283 (DRAK1 IC_{50}: 10−30 nM) and the TBK1/IKKε inhibitor BAY-985 (DRAK1 IC_{50}: 311 nM).[27,28] Recently, Picado et al. have developed a chemical probe for DRAK2 (IC_{50}: 34 nM) with good selectivity against the kinome and weak activity on the related DRAK1 isoform (IC_{50}: 4700 nM).[29] The selectivity between the DRAK1 and DRAK2 isoforms has been surprising, considering the high degree of sequence conservation.
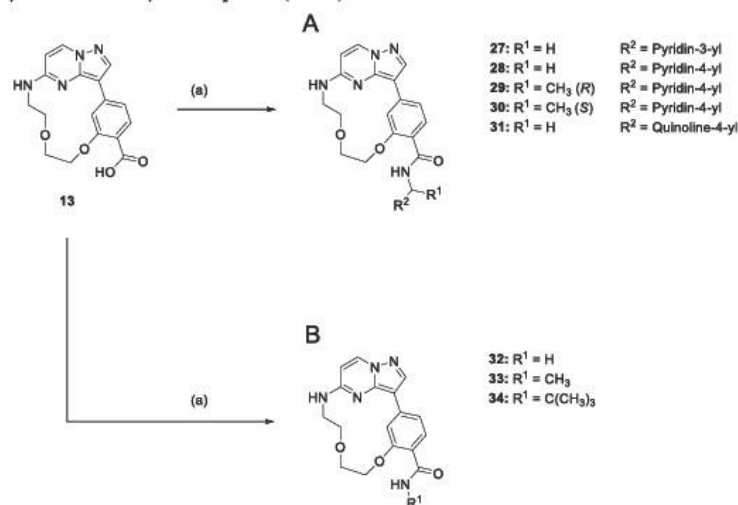
A versatile pharmacophore that we already utilized for the development of potent kinase inhibitors are pyrazolo[1,5-a]pyrimidines.[30] This versatile hinge binding moiety has also

been used for the development of macrocyclic inhibitors targeting RIPK2 and ALK2, and we used this scaffold recently for the development of a highly selective inhibitor targeting the kinase CK2.[31−33] These studies demonstrated how small modifications in macrocyclic inhibitors can result in significant differences in selectivity. With this study, we report the structure-based optimization of the CK2 inhibitor IC19[31] resulting in an exclusively selective DRAK1 inhibitor CK156 (34) via an optimization of back pocket interactions (Figure 1). The cellular activity on DRAK1 was confirmed to be in the nanomolar range by a NanoBRET target engagement assay and a high-content screen revealed only a minor cytotoxicity (>10 μM) in four different tested cell lines. Surprisingly, studies in GBM cell lines showed that selective targeting of DRAK1 kinase activity in this cancer did not result in the expected cellular phenotype and toxicity reported by DRAK1 knockdown experiments suggesting that the catalytic function of DRAK1 is not required for cell survival.[25]

2. RESULTS AND DISCUSSION

2.1. Identification and Verification of Pyrazolo[1,5-a]pyrimidine-Based Macrocycles as DRAK1 Inhibitors. The starting point for the development of a series of selective inhibitors targeting DRAK1 was the non-selective macrocyclic inhibitor IC19, which was based on the well-established pyrazolo[1,5-a]pyrimidine scaffold.[30,34] In our previous work, we have shown that the carboxylic acid in IC19 was crucial for CK2 activity. Initial modifications of the carboxylic acid motif toward the back pocket, via amide coupling using N-benzylamine resulting in the N-benzylamide derivative (14) revealed significant activity of compounds for DRAK1 in

Scheme 2. Synthesis of Macrocyclic Compounds (27–34)[a]

27: R¹ = H		R² = Pyridin-3-yl
28: R¹ = H		R² = Pyridin-4-yl
29: R¹ = CH₃ (R)		R² = Pyridin-4-yl
30: R¹ = CH₃ (S)		R² = Pyridin-4-yl
31: R¹ = H		R² = Quinoline-4-yl

32: R¹ = H	
33: R¹ = CH₃	
34: R¹ = C(CH₃)₃	

[a](A) Heteroaromatic amide derivatives and (B) small amide derivatives. Reagents and general conditions: (a) (1) HATU, DIPEA, dry DMF, argon atmosphere, 1 h, rt. (2) Corresponding amine, 18 h, rt.

differential scanning fluorimetry (DSF) assays. Even though the initial compound (14) was not selective for DRAK1, it inhibited only few off-targets and showed no activity against the closely related family member DRAK2 and DAPK3 another member of the DAPK family (Figures 1 and 3). In order to guide further optimization of 14, we solved the structure of DRAK1 in complex with 14 (Figure 2). To our knowledge, this represents the first crystal structure of the catalytic domain of this kinase. The structure of DRAK1 showed high structural similarity with DRAK2. Comparison with the DRAK2 quercetin complex (pdb-ID: 3LMN) revealed similar structural elements including the unusual, non-conserved two small beta sheets located N-terminal to helix αC. Interestingly, the structure of DRAK1 revealed an activation loop exchange with a second kinase domain located in the asymmetric unit. This unusual conformation has been described for kinases requiring auto-activation at non-consensus sites located in the activation segment including the DAPK family member DAPK3.[35,36] Indeed, in the observed conformation, T226 is positioned in the substrate-binding site of the interacting protomer. However, whether DRAK1 requires phosphorylation at this position for full catalytic activation is not known. T226 is conserved in DRAK2 (T198). In the active conformation of the activation segment, a phosphorylation at T226 would stabilize an interaction with a lysine residue (K188) in the catalytic loop following the HRD (HLD in DRAK1) motif resulting in the stabilization of the substrate-binding site.

The binding mode of 14 revealed that it interacted with the ATP binding pocket of DRAK as a canonical type I inhibitor, which was anchored via the pyrazole nitrogen to the backbone nitrogen of A141 located in the kinase hinge region. An additional hydrogen bond was formed by the carbonyl oxygen

to the conserved back pocket lysine (K90) in the β3-strand. A surface representation of DRAK1 co-crystallized with 14 revealed a large unoccupied back pocket, to which the benzyl moiety of 14 was pointing. A structural alignment of the DAPKs confirmed that DAPK1–3 differed significantly from DRAK1 and DRAK2, whereas the ATP-binding pocket of DRAK1 shared high similarity with DRAK2.

2.2. Synthesis of Macrocyclic Pyrazolo[1,5-a]-pyrimidines. Based on the structure of DRAK1 with 14, further efforts were made to explore the back pocket and a series of macrocycles were synthesized, by focusing on the derivatization of the carboxylic acid of IC19. Modifications were introduced via amide coupling using HATU under basic conditions in DMF. The synthesis of the macrocyclic core scaffold 13 progressed by employing the synthesis route already established and described[31] (Scheme S1). In the first series of our optimization, various benzylamide derivatives were introduced (15–21). These molecules were substituted at the 2 positions of the pendant ring system by introducing chlorine, fluorine, and methyl groups (Scheme 1A). A naphthalen-1-ylmethanamide (22), a 1-(naphthalen-1-yl)-ethanamide (26) as well as 1-phenylethanamides (23–25) with different stereoisomerism were realized in a second series of DRAK1 inhibitors in order to diversify the inhibitor more at this key position (Scheme 1B). In a third series, hetero-aromatic derivatives were synthesized (Scheme 2A). For this purpose, pyridin-3-ylmethanamines, pyridin-4-ylmethan-amines, 1-(pyridin-4-yl)ethanamines (R and S conformation), and quinolin-4-ylmethanamines were reacted via an amide coupling at the carboxylic acid to obtain the corresponding compounds 27–31. Small non-aromatic amides were intro-duced in the fourth series. For this, ammonia, methanamine, and 2-methylpropan-2-amine were used as educts. The

Scheme 3. Synthesis of Macrocyclic Compounds with Heterocyclic Residues (35 and 36)[a]
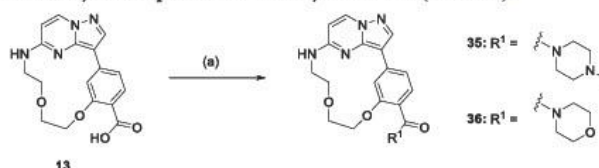
[a]Reagents and general conditions: (a) (1) HATU, DIPEA, dry DMF, argon atmosphere, 1 h, rt. (2) Corresponding amine, 18 h, rt.

resulting compounds 32–34 are shown in Scheme 2B. In the last series, heterocyclic amides were introduced via amide coupling and compounds 35–36 were obtained (Scheme 3).

2.3. Structure–Activity Relationship of Pyrazolo[1,5-a]pyrimidine Macrocycles. To assess the binding and selectivity of the synthesized compounds against DRAK1 and potential off-targets, a DSF assay was performed.[37] Using this sensitive assay that determines target protein stabilization by inhibitor binding via a shift of the melting curve (higher shift corresponds to higher stabilization and hence higher affinity), the macrocycles were tested against DRAK1 as well as against a panel of 44 diverse kinases (final protein concentration: 2 μM; final inhibitor concentration 10 μM) in order to assess initial selectivity of the synthesized compounds against the kinome (Figure 3/Table S1). In a first SAR-series, the substitution pattern at the benzyl moiety was explored by the introduction of mainly halogenated residues and tolyl residues and the ΔT_m shifts were compared to parental compound 14 during the complete SAR. Only the double chlorinated derivatives in the ortho and meta positions (20) led to small ΔT_m shifts suggesting loss of activity against DRAK1, whereas 16, 18, and 21 showed about the same shift, and 15, 17, and 19 lead to higher shifts. The selectivity of 15 showed improvement in selectivity against our screening panel with just 4 kinases identified with ΔT_m shifts larger than 5 °C, but the potent DRAK1 inhibitors 17 and 19 showed broad activity interacting strongly with 6 and 12 kinases, respectively.

In a second SAR-series, we introduced bulkier aromatic residues, such as the 1-napthyl derivative (22) but also more branched moieties (23–26), by the introduction of an additional methyl group in the benzylic position. Whereas both 1-naphthyl derivatives (22 and 26) showed smaller ΔT_m shifts for DRAK1, the introduction of the additional methyl group in benzylic position 23–25 (rac, S, R) resulted in a significant increase in ΔT_m shifts. However, in particular, 23 (rac) and the (R)-enantiomer (25) were less selective in this kinase panel.

In a third optimization step, more heteroaromatic derivatives were synthesized (27–31). Increased ΔT_m shifts were identified for the pyridine derivatives 27–30. Again, the bulkier quinoline (31) was not tolerated. This confirmed the trend already observed for 20, 22, and 26 revealing that large asymmetric residuals were not preferred at this position. In a fourth SAR-series, we removed the benzyl moiety in 14 and introduced smaller motifs at the amide (32–34). The switch to small amidic residues (32–34) revealed a significant increase in binding affinity. However, small amide 32 was the least selective derivative of this series likewise the N-methyl amide (33) interacted with a significant number of additional kinases. Intriguingly, the bulkier tert-butyl moiety in 34 had excellent selectivity with just 2 other kinases that showed

significant ΔT_m shifts. Therefore, all in all for compound 34, just 3 kinases have been identified with ΔT_m shifts ≥5 °C (namely, DRAK1, BIKE, and AAK1; Table S1). In a last optimization attempt, the heterocyclic derivatives 35 and 36 were synthesized to provide alternative interactions with the back pocket. However, the introduction of the heterocyclic derivatives (35 and 36) dramatically decreased the affinity of these macrocycles toward DRAK1 and most kinases in our DSF panel. These very low affinities toward DRAK1 provided a potential use of 35 and 36 as negative control compounds. In our DSF selectivity panel, the NAK kinases (AAK1, BIKE, and GAK) were identified as potential off-targets for most compounds of this series. Compared to 14, the affinity toward aurora-B was reduced by introducing substituted benzylamide derivatives. Intriguingly, none of the compounds showed a significant shift for the closely related DRAK2. In conclusion, the established SAR series revealed that modifications at moieties interacting with the back pocket increased both binding affinity for DRAK1 and selectivity within the chosen selectivity panel. Compounds 23, 30, and 34 achieved the highest ΔT_m shifts for DRAK1. The introduction of the methyl group in the benzylic position (23) as well as in the pyridine derivative (30) led to a similar increase in binding affinity as judged by ΔT_m shifts as observed for the highly selective tert-butyl derivative (34) (Figures 3 and 4).

2.4. CK156 (34) and CK228 (30) Show Excellent Potency for DRAK1. In order to determine the affinity of the two most promising DRAK1 inhibitors CK228 (30) and CK156 (34) in solution, we used isothermal titration calorimetry (ITC) (Figure 5). Both compounds revealed low K_D values of 9 nM (30) and 21 nM (34), respectively. These data were in good agreement with the ΔT_m-shift data that showed shifts of 12.3 and 11.3 °C, respectively (Table 1), that reflects a higher ΔT_m shift leads to lower K_D values. An interesting aspect of ITC data is that the binding thermodynamics of the inhibitor interaction can be assessed. The enthalpic contributions to the binding were comparable for 30 and CK156 (34) showing large negative enthalpy changes upon binding indicative of favorable polar interactions. Intriguingly, the changes in entropy upon binding ($T\Delta S$) were also negative, contributing to binding. These data suggested that macrocyclization of these compounds reduced compound flexibility and most likely stabilized a conformation close to the bioactive conformation.

We tested compounds 14, 23, 30, and CK156 (34) in [33]PanQinase enzyme kinetic assays provided by ProQinase (reaction biology Europe; Table S11). This commercial assay format represents a plate-based variant of a scintillation proximity assay (SPA). In this radiometric activity assay, a ^{33}P-labeled phosphate of ATP is transferred from the kinase to
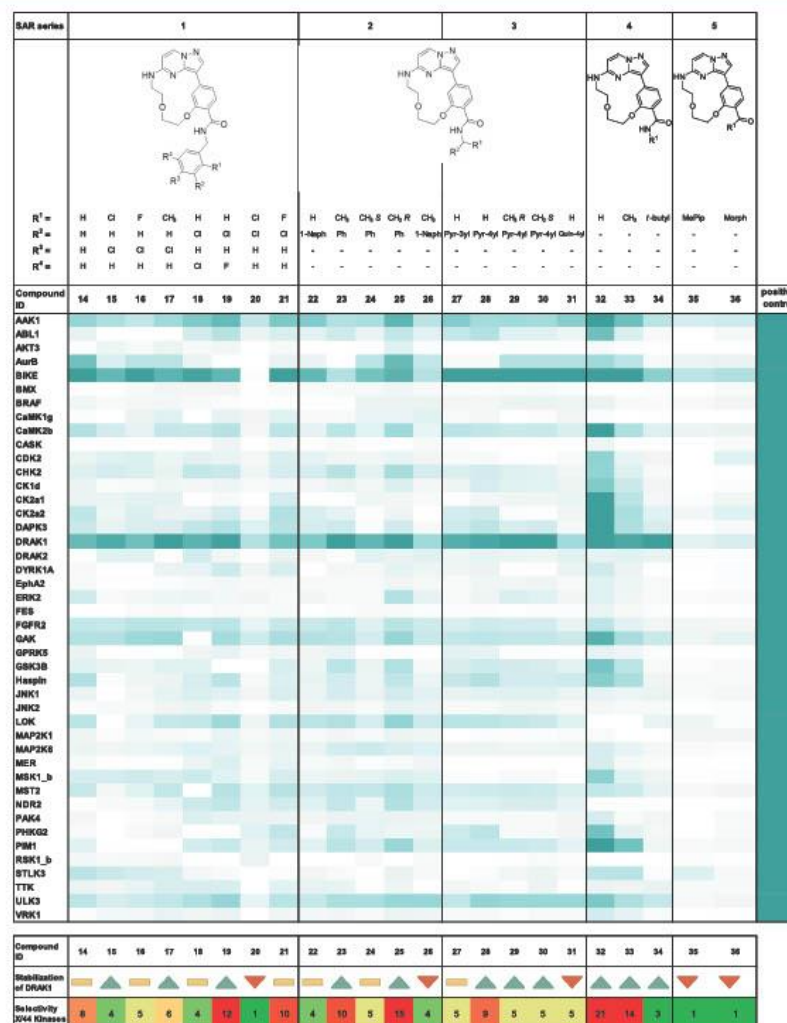
178

Figure 3. Structure–activity relationship and selectivity of macrocyclic pyrazolo[1,5-a]pyrimidines. In a DSF assay, compounds 14–36 were screened against a panel of 44 protein kinases using staurosporine as a positive control (final protein concentration: 2 μM; final inhibitor concentration 10 μM). The results are presented as a heat map. Kinases that are not stabilized by the compounds are colored white and kinases with the highest stabilization are colored in dark green—where the color for each line is normalized against the respective staurosporine shift. The graphic below the ΔT_m shift panel shows temperature shifts compared to the lead structure 14 (green improvement, yellow unchanged, and red less favorable selectivity). Also, the selectivity within the panel of 44 kinases, is shown, respectively, which corresponds to the number of kinases with a ΔT_m-shift > 5 °C. The values of the ΔT_m shifts are listed in Table S1.

the kinase substrate.[38,39] The resulting IC$_{50}$ values confirmed the potency of the inhibitors for DRAK1. The lead compound of our optimization 14 showed an IC$_{50}$ value of 155 nM, whereas inhibitor potency was significantly improved for 23 and 30 with IC$_{50}$ values of 31 and 15 nM, respectively. The potency of CK156 (34) was slightly below that of 23 and 30

179
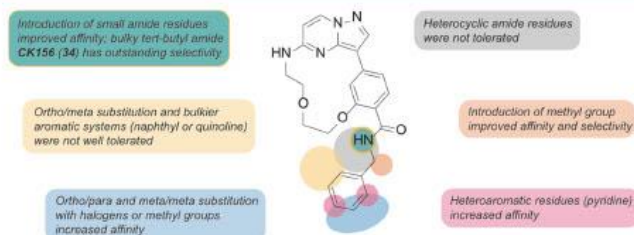
Figure 4. Structure–activity relationship summary of the modifications carried out using 14 as a chemical starting point.
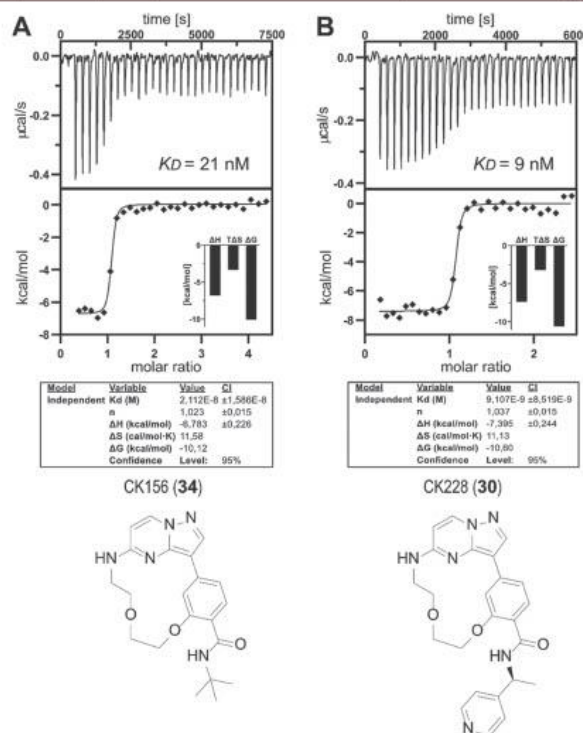
Figure 5. Binding affinity of CK156 (34, panel A) and CK228 (30, panel B) determined by ITC revealed low nanomolar binding affinity of these compounds for DRAK1. The chemical structure of the two compounds is shown in the lower panel.

but it was still in the low nanomolar range with an IC_{50} of 49 nM. Compared to the ITC experiments, the data in the ^{33}PanQinase assay (by ProQinase) were in good agreement with determined binding constants as well as with the DSF data (Table 1).

2.5. CK156 (34) Is a Highly Selective DRAK1 Inhibitor. After having obtained promising selectivity data for compounds 23, 30, and 34 in our DSF selectivity panel, their

selectivity profile was determined in the kinome-wide KINOMEscan assay by Eurofins. For comparison of the selectivity profiles, the lead structure 13 and the parental compound 14 was also screened for this SAR study. The scanMAX panel covers 80% of the human kinome and included 468 kinases. The compounds were screened at two concentrations (0.1 and 1 μM) except 30, which was only screened at 1 μM (Figure 6 and Tables S2–S9). The

Table 1. Binding Affinities of the Promising Compounds 14, 23, 30, and 34–36 against DRAK1 from Various Assay Platforms (DSF, ITC, ^{33}PanQinase, and NanoBRET) and Affinities against the Main Off-Targets CK2α1 and CK2α2 Identified by KINOMEscan^{a}

	DSF T_m-shift			ITC K_D	^{33}PanQinase IC$_{50}$			NanoBRET IC$_{50}$			KINOMEscan S_{35}
compound	DRAK1	CK2a1	CK2a2	DRAK1	DRAK1	CK2a1	CK2a2	DRAK1	CK2a1	CK2a2	100 nM/1 μM
13	7.9	10.6	11.3								-/0.26
14	9.3	1.7	4.9		155 nM			171 nM			0.03/0.09
23	12.1	−0.9	1.9		31 nM			71 nM			0.02/0.05
30	12.3	1.2	0.1	9 nM	15 nM			70 nM			-/0.18
34	11.3	0.8	2.3	21 nM	49 nM	950 nM	380 nM	182 nM	34 μM	39 μM	0.005/0.02
35	2.1	0.3	0.9					>45 μM			
36	3.0	1.1	2.5					>45 μM			

^{a}Compared to the macrocyclic ground structure 13, all compounds showed an improved affinity towards DRAK1 in the DSF assay. The potencies in vitro are in a range of low two-digit and low three-digit nanomolar. The cellular affinities are in a comparable range. All compounds show no significant off-target affinities.

quantification of the selectivity of the compounds was based on the selectivity score (S-score) calculated by the ratio of kinase hits at a set cut off by the total number of tested kinases. A value of 35% displacement (35% of Ctrl) was used as a potency threshold resulting in S_{35} = (number of non-mutant kinases with %Ctrl <35)/(number of non-mutant kinases tested).[40] Whereas lead structure 13 also revealed potent inhibition of DRAK1, it was highly unselective in the KINOMEscan assay at 1 μM (S_{35} = 0.26). Compounds 14 and 23 already showed a very low selectivity score (S_{35} = 0.03 and S_{35} = 0.02) at a screening concentration of 0.1 μM. At an increased concentration (1 μM), 14 (S_{35} = 0.09) bound to 41 kinases (Table S4), which were almost twice as many as 23 (S_{35} = 0.05) (Table S6). Gratifyingly, compound 34 showed an even better selectivity at both 0.1 μM (S_{35} = 0.005) and 1 μM (S_{35} = 0.02). Surprisingly, the selectivity score of 30 at 1 μM was S_{35} = 0.18, higher by a factor of 10 than that of 34, and 65 kinases were bound. To validate the encouraging data of 23 and 34, an enzymatic activity ^{33}PanQinase (ProQinase) was performed. The activity of the off-targets CK2a1 and CK2a2, detected in the KINOMEscan assay was 20-fold and 10-fold weaker, respectively, than IC$_{50}$ measured for DRAK1 (Table 1). These data were in agreement with our DSF assay that showed only weak ΔT_m shifts for CK2. Intriguingly, from the other two off-targets identified in our DSF-Panel, namely, AAK1 and BIKE (BMP2K), just the last-mentioned appeared only at the higher screening concentration of 1 μM but not at the lower concentration at 0.1 μM. Thus, 34 showed an excellent selectivity profile in two comprehensive binding and enzymatic assay panels, apart from very weak activity on the off-targets GAK and BMP2K and the closely related DRAK2 at a screening concentration of 1 μM.

Supporting the high selectivity in kinome-wide screens, the crystal structure of compound CK156 (34) revealed excellent shape complementarity with the DRAK1 active site (Figure 7A). As expected, similar to the benzyl residue of 14, the tert-butyl residue of CK156 (34) was pointing toward the back pocket of DRAK1 and hydrogen bonds with the hinge region and lysine K90 were conserved.

The alignment of DRAK1 co-crystallized with CK156 (34) and 14 revealed interesting features of the binding mode of CK156 (34). It seems that due to the smaller tert-butyl moiety (34) in comparison to the benzyl moiety (14), CK156 (34) can bind deeper to the hinge region (see Figure 7B). In addition, a slight conformational change of the P-loop was noticed in the case of 14 (Figure 7C). However, this deeper

binding into the hinge region was only possible due to a conformational rearrangement in the hinge region. The hinge region of DRAK1 and the subsequent linker contains a very small amino acid sequence (AAGG; see Figure 2D). Especially, glycine 143 (G143) was twisted in the binding mode of CK156 (34) in comparison to the binding mode of 14 (see Figure 7B). This flexibility within the hinge region is unusual for a kinase and just a few examples have been described in the literature, for example, for Skepinone-L, a highly selective p38 inhibitor,[41,42] where a complete glycine flip in the hinge region has been observed. This was not observed in this manner for CK156 (34) and may indicate a level of flexibility in the back pocket. As summarized, we have seen that DRAK1 covers some flexible regions in the ATP-binding pocket that might explain the promiscuity of DRAK1 to other kinase inhibitors.

2.6. NanoBRET Confirmed CK156 (34) as a Highly Selective and Potent Inhibitor in Cells. Due to the promising kinome-wide selectivity data and the low nanomolar potency, the cellular activity of 23 and 34 was assessed in the NanoBRET target engagement assays. This method represents a tracer displacement assay measured in live cell systems. Hence, the quantitative analysis of compound binding is performed under physiological ATP concentrations using full-length kinases and allows the determination of IC$_{50}$ values in cellular environments. We tested compounds 14, 23, 30, 34–36, and staurosporine as a positive control in the DRAK1 NanoBRET assay. The IC$_{50}$ values for 23 and 30 were similar to 70 and 71 nM, respectively, and 14 and 34 showed IC$_{50}$ values of 171 and 182 nM, respectively. Compared to the values determined with ITC and ^{33}PanQinase assays, there was a marginal drop in potency between enzymatic assay data and cellular NanoBRET values suggesting excellent cell penetration of these inhibitors (Table 1). Staurosporine showed an IC$_{50}$ value of 54 nM, which was comparable to the published in vitro data.[43] To evaluate the possible off-targets identified by KINOMEscan, further NanoBRET assays with CK2a1, CK2a2, GAK, BMP2K, and DRAK2 were performed. The IC$_{50}$ values were more than 30-fold higher compared to potencies measured for 34 against DRAK1 (Figure 8). As expected, 35 and 36 showed no significance in cellulo activity on DRAK1 (IC$_{50}$ > 45 μM) and are predestined as the corresponding negative control compounds for 34 (CK156).

2.7. Cellular Effects and Cytotoxicity of DRAK1 Inhibition. In CK156 (34), we have identified the first potent and selective DRAK1 inhibitor. Despite the fact, that DRAK1 belongs to the understudied kinases, some potential biological
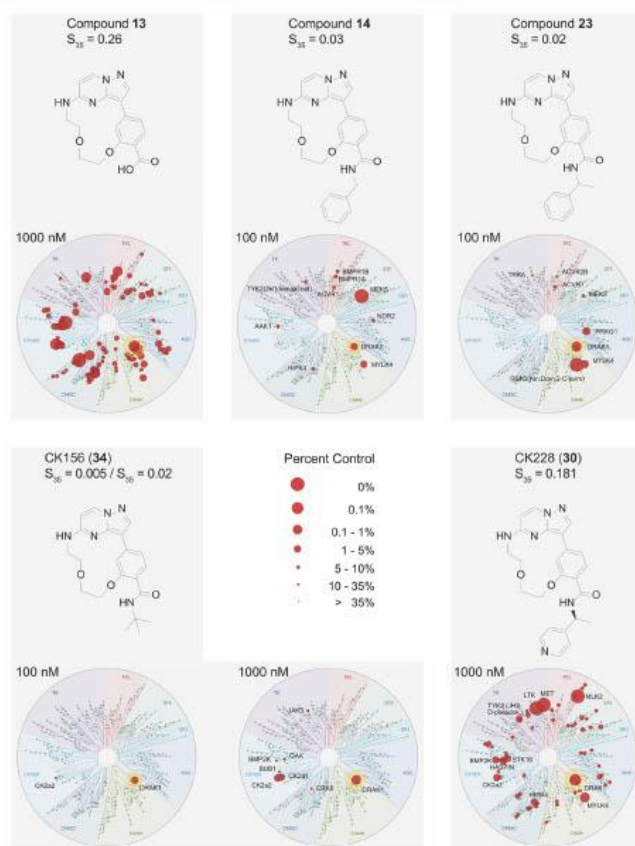
Figure 6. Selectivity profile of pyrazolo[1,5-*a*]pyrimidine-based macrocycles. Shown are selectivity profiles for the non-selective macrocyclic core structure 13, the lead structure 14 as well as the compounds 23, CK156 (34), and CK288 (30). The panel included 468 kinases and displacement data were measured in the KINOMEscan assay (Eurofins) at 0.1 and 1 μM, respectively. The red circles represent the displacement of the control and their size corresponds to the % displacement (small value corresponds to high potency) as indicated in the figure. The S-score quantifies the selectivity at the measured concentration; threshold of %Ctrl = 35% [S_{35} = (number of non-mutant kinases with %Ctrl <35)/(number of non-mutant kinases tested)].

effects have been attributed to DRAK1 in the literature. DRAK1 has been described to be highly overexpressed in gliomas compared to the normal brain tissue and other cancer cells.[25] Knockdown of DRAK1 resulted in an alteration of the cell shape that is associated with a decrease in proliferation, clonogenicity, and migration. In addition, it was demonstrated that DRAK1 knockdown also sensitized GBM cells to DNA damaging agents.[25]

To assess the efficacy of these highly selective macrocyclic DRAK1 inhibitors toward glioma cells, we first investigated their effect on general cell viability using a multiplex high-content assay simultaneously allowing the observation of phenotypic effects (Figure 9). CK156 (34) was screened as the

compound with the best selectivity and CK228 (30), the macrocycle with the highest potency toward DRAK1, against three different cell lines, human embryonic kidney cells (HEK293T), osteosarcoma cells (U2OS), and non-transformed human lung fibroblasts (MRC-9) at two different concentrations (1 and 10 μM) (Figure 9A). Cells were treated with fluorescent dyes to detect apoptosis (Alexa Fluor 680 conjugate), nuclei/DNA (Hoechst33342), mitochondria function (Mito Tracker red), and microtubule effect (BioTracker 488 green microtubule cytoskeleton dye). Fluorescence and cellular shape were measured before compound treatment and 18 h (Figure 9C) as well as 36 h after exposure to the compounds [CK156 (34), CK228 (30)], respectively, using
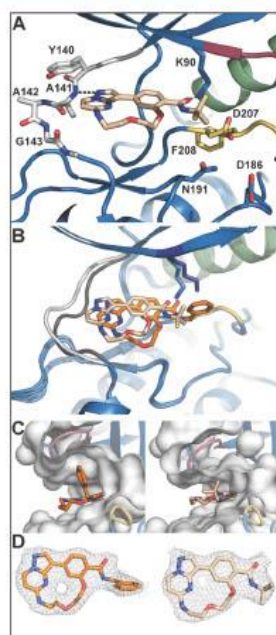
Figure 7. Binding mode of CK156 (34) (A), alignment of 14 and 34 (B), view into the catalytic pocket containing 14 and 34 from direction of the back pocket (C), and electron density of 14 and 34 (D). The binding mode of the beige stick represented 34 in DRAK1 (PDB: 7QUF) and is also ATP competitive (type I) and the pyrazolo[1,5-a]pyrimidine interacts with the hinge via the backbone nitrogen of A141. The carbonyl oxygen of 34 shows the same interaction with the side chain of K90 as 14. Alignment of the binding modes of 14 and 34 shows that 14 binds slightly deeper in the catalytic pocket, resulting in a slight deformation of the hinge. In (C), it can be seen that, compared to the smaller tert-butyl residue of 34, there is a slight displacement of the P-loop when 14 binds with the space-filling aromatic system.

the CQ1 high-content confocal microscope (Yokogawa). Staurosporine at a concentration of 10 μM was used as a reference compound.

There were no significant cytotoxic effects apparent after 18 h of compound treatment with 1 μM of CK156 (34) or CK228 (30), respectively, and the distribution of healthy versus dying cells was similar for CK156 (34) and CK228 (30), respectively, compared to 0.1% DMSO control (Figure 9B). At the higher concentration, a small increase in the number of apoptotic cells was observed and the more potent compound CK228 (30) showed a slightly lower cytotoxic effect after 18 h at 10 μM compared to CK156 (34). After 36 h no cytotoxic effects were observed at the lower concentration (1 μM) as well (Figure 9A). Treatment of cells with 10 μM for 36 h resulted in a slightly higher apoptotic rate for both compounds shown in all three cell lines. However, the EC$_{50}$ in all cells were higher than 10 μM except for CK156 (34) in the human osteosarcoma cells (U2OS) and no notable effect on microtubule or mitochondrial damage was observed with any of the compounds in all cell lines used. To sum up, at this point, we have seen that these compounds [CK156 (34) and CK228 (30)] showed no general cytotoxicity at 1 μM in the tested cell lines, whereas at 10 μM some toxicity was observed. However, because both compounds were not tested in the KINOMEscan at 10 μM, we cannot completely rule out whether these are some unspecific cytotoxic effects or if these effects results from the inhibition of other kinases at the higher concentration. To study the efficacy of these macrocycles in glioma cells, we first analyzed the expression of DRAK1 in five different GBM cell lines (Figure S2). The cell lines LN-229 and D-247MG had the highest expression levels of DRAK1, whereas T98G and LN-18 showed low expression levels. Interestingly, U-87-MG, which has been shown to be sensitive to DRAK1 depletion,[25] also expressed relatively low amounts of this kinase. Cell viability assays of these cell lines exposed to CK156 (34) and CK228 (30) at a concentration range of 0–20 μM in a dose-dependent manner were performed and analyzed after 1 and 7 days of compound treatment, respectively. The EC$_{50}$ values for CK156 (34) were very close (9–11 μM), except for the activity toward DRAK1 in D-247MG cells (Figure 9D). Here, the EC$_{50}$ value obtained was 2.8 μM. Treatment of the cell lines with CK228 (30) resulted in EC$_{50}$ values between 1 and 2 μM (Figure S3A). Surprisingly, these data did not correlate with the low nanomolar activity of compounds 34 and 30 against DRAK1 in the NanoBRET assay. We also did not observe any alteration in the cell shape
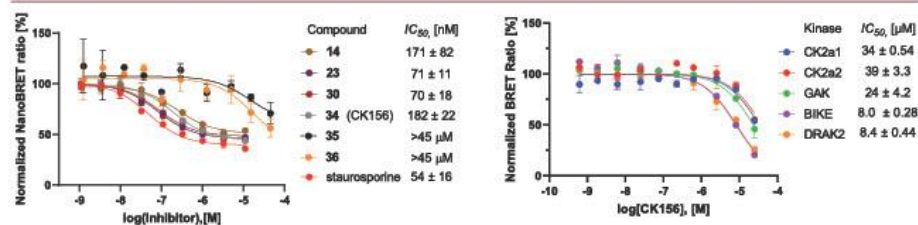
Figure 8. Cellular potencies of macrocyclic pyrazolo[1,5-a]pyrimidines on DRAK1 and of CK156 (34) on potential off-targets CK2a1, CK2a2, GAK, BIKE, and DRAK2 determined by NanoBRET in HEK293T cells. The respective IC$_{50}$ values and errors are given in the figure legend. The data represents measurements of $n = 4$ for inhibitors profiled on DRAK1, except the negative control candidates 35 and 36 as $n = 2$, as well as $n = 2$ for the off-target profiling.
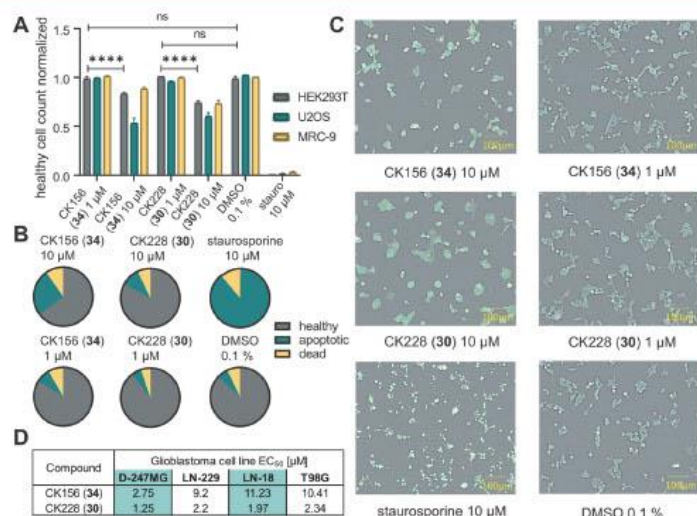
Figure 9. Live cell high content screen of HEK293T, U2OS, and MRC-9 cells. (A) Healthy cell count after 36 h of 10 and 1 μM compound exposure [CK156 (34), CK228 (30), and staurosporine] normalized to healthy cells exposed to DMSO (0.1%) in HEK293T, U2OS, and MRC-9 cells. Error bars show standard error of the mean of quadruplicates. Significance was calculated using a two-way ANOVA analysis. (B) Fraction of healthy, apoptotic and dead cells after 18 h of 10 and 1 μM compound exposure [CK156 (34), CK228 (30), and staurosporine] in comparison to the 0.1% DMSO control in HEK293T cells. (C) Brightfield confocal image of stained HEK293T cells after 18 h of 10 and 1 μM compound exposure [CK156 (34), CK228 (30), staurosporine] in comparison to 0.1% DMSO control (blue: DNA/nuclei, green: microtubule, red: mitochondria, and magenta: Annexin V apoptosis marker). Additional high content data shown in supplementary Figure S1. (D) EC$_{50}$ values of CK156 (34) and CK228 (30) in glioblastoma cell lines (D-247MG, LN-229, LN-18, and T98G) after 7 d analyzed with resazurin.

as described in knockdown experiments of DRAK1 in U87, SF268, A172, and U118 cells.[25] Only minor changes in the morphology were observed in the low expressing T98G cells (Figure S3B). It therefore seems that an inhibition of the kinase activity alone does not replicate the observed phenotype in GBM.

3. CONCLUSIONS

In this study, we demonstrated that modifications in a pyrazolo[1,5-a]pyrimidine-based macrocyclic core scaffold resulted in the development of a highly selective DRAK1 inhibitor. The introduction of space-filling benzylamide derivatives interacting with the back pocket significantly improved the kinome wide selectivity (KINOMEscan). Further optimization resulted in highly potent inhibitors (23, 30) of DRAK1 with in vitro IC$_{50}$ values of 31 and 15 nM, respectively. The influence of derivatization with small amides on kinase inhibition was also investigated. Here, CK156 (34) entered the spotlight as a potential probe candidate. Although its potency with an in vitro IC$_{50}$ value of 49 nM is somewhat weaker than those of 23 and 30, it is still well below 100 nM, which was considered the criteria for a high-quality chemical probe.[44] CK156 (34) demonstrated remarkable selectivity in KINOMEscan with S_{35} = 0.005 (100 nM) and S_{35} = 0.02 (1 μM), respectively, and showed exclusive affinity to DRAK1. Potential off-targets showed a 30-fold lower affinity for CK156 (34) in cellulo compared to DRAK1 with an IC$_{50}$ value of 182 nM. Crystallization experiments resulted in the first structures

of DRAK1 and revealed a classical type I inhibitor binding mode, as expected. In cell viability assays, we observed that CK156 (34) can be utilized as a chemical tool compound at a concentration of 1 μM without any cytotoxic effect in HEK293T, U2OS, and MRC-9 cells. Only in U2OS cells, we have observed some unspecific toxicity at a higher concentration of 10 μM. Previous experiments by Mao et al. demonstrated that the knockout of DRAK1 resulted in dramatic alteration in the cell shape.[25] This effect was associated with decreased proliferation, clonogenicity, migration, invasion, and anchorage independent colony formation. However, our selective tool compound CK156 (34) revealed an EC$_{50}$ below 10 μM and only showed an effect in one cell line, D-247MG, which displayed the highest DRAK1 expression level (EC$_{50}$ = 2.75 μM). However, no correlation between the DRAK1 expression and the effect on GBM viability or morphological changes was observed. Thus, it is likely that scaffolding effects rather than kinase activity are responsible for the phenotype observed in knockdown studies. Stronger effects were observed for the more potent inhibitor CK228 (30) (EC$_{50}$ < 3 μM) in all four GBM cell lines tested. However, it is likely that these effects are not solely due to on-target effects by DRAK1 inhibition but that they may be related to off-target effects, due to the lower selectivity of this inhibitor (S_{35} = 0.18@1 μM).

Taken together, we have identified with CK156 (34) a chemical probe for DRAK1 together with a suitable negative control compound (36). In previous work, it has been reported

that DRAK1 is a novel candidate for therapeutic targeting in glioblastoma. However, our initial cellular results with this chemical probe for DRAK1 revealed discrepancy between the reported knockdown experience and chemical inhibition in four different GBM cell lines. In future studies, it would be useful to evaluate more selective DRAK1 inhibitors with different chemical scaffolds to identify if these results were compound-related or target-related effects of DRAK1 inhibition. Given these insights, it is evident that further systematic investigation of the essentiality of DRAK1 in glioblastoma cell lines is warranted to clarify the role of DRAK1 in glioma and to define the potential for the prospective therapeutic application of selective DRAK1 inhibition. In addition, it would be worthwhile to investigate the effect of inhibition of DRAK1 by small molecules also in other cancer diseases to assess the therapeutic potential of DRAK1. We believe that our selective chemical probe CK156 (34) will help to elucidate the biological role of DRAK1.

4. EXPERIMENTAL SECTION

4.1. Cloning and Expression. DNA encoding for human DRAK1 residues V50—P350 was cloned into the kanamycin resistant expression vector pNIC28-Bsa4 that includes a N-terminal His$_6$-tag and a TEV protease cleavage site. DNA coding for residues A39—E369 was cloned by ligation-independent cloning into expression vector pNIC-H102 that includes a N-terminal His$_{10}$-tag, a TEV protease cleavage site, and a kanamycin resistance gene. For the expression of His-tagged DRAK1, the plasmid was transformed into competent Rosetta (DE3) *Escherichia coli* cells (Novagen). Cultures were grown in a terrific broth medium at 37 °C until an OD$_{600}$ of 1 was reached. The temperature was reduced to 18 °C and after 30 min of incubation, the protein expression was induced using IPTG. Cultures incubated overnight at 18 °C to allow for sufficient protein expression. The cells were then harvested by centrifugation (6000 rpm, 15 min, 4 °C).

4.2. Purification. For the purification of DRAK1$_{KD}$, bacterial cell pellets were resuspended in lysis buffer (50 mM HEPES pH 7.4, 500 mM NaCl, 20 mM imidazole, 0.5 mM TCEP, and 5% glycerol) and lysed by sonication (35% amplitude, 5 s pulse and 10 s pause, and 12 min total pulse time). After centrifugation (23,000 rpm, 30 min, 4 °C) to clear the lysate, the supernatant was loaded onto a Ni NTA column. The column was washed with lysis buffer (150 mL) and His-tagged protein eluted with lysis buffer containing 300 mM imidazole. The protein was dialyzed overnight in 2 L dialysis buffer (50 mM HEPES pH 7.4, 500 mM NaCl, 0.5 mM TCEP, and 5% glycerol) and the His$_6$-tag cleaved by adding TEV protease. The next day, the cleaved Tag and the TEV protease were removed by loading the protein solution onto a Ni NTA column. The flowthrough containing the protein of interest was concentrated and subjected to gel filtration using a HiLoad 16/600 Superdex 200 pg gel filtration column connected to an AKTA Xpress system with gel filtration buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 0.5 mM TCEP, and 5% glycerol). Protein containing fractions were pooled together and concentrated to approximately 3 mg/mL.

4.3. Crystallization. DRAK1 (residues 50—350) was co-crystallized with 14 and 34, respectively, using the sitting-drop vapor diffusion method. 200 nL sized drops, containing the protein solution and reservoir solution in ratios of 2:1, 1:1, and 1:2, were set up on a 3-well crystallization plate (SwissCI). For the co-crystallization experiment of DRAK1 with 14, crystals were obtained after mixing the protein at a concentration of 15 mg/mL with the precipitant solution containing 17% PEG3350 and 0.2 M ammonium formate in a ratio of 2:1. Crystals of DRAK1 with 34 (CK156) were obtained after mixing the protein at a concentration of 9 mg/mL with the precipitant solution containing 24% PEG3350 and 0.2 M ammonium chloride in a ratio of 1:2 of protein to the precipitant solution.

Crystals grew at 4 °C within days and were mounted in precipitant solution containing additional 25% ethylene glycol for cryoprotection.

4.4. Data Collection and Structure Determination. Diffraction data were collected at swiss light source beamline X06SA (Villigen, CH) at a wavelength of 1.0 Å, analyzed, scaled, and merged with the SLS automated data processing (adp) pipeline.[45] Structures were solved by molecular replacement using Phaser[46] with STK17B (PDB ID: 3LM0) as a model. The structure models was built using Coot[47] and refined using Refmac5.[48] A summary of the crystallographic data collection and refinement data can be found in Table S12. The models of DRAK1 with 14 and 34 have been deposited to the PDB and can be found as PDB IDs 7QUE and 7QUF.

4.5. Differential Scanning Fluorimetry Assay. Thermal melting experiments were carried out as described previously[31] using a Mx3005p real-time PCR machine (Agilent). Proteins were buffered in 25 mM HEPES (pH 7.5) and 500 mM NaCl and were assayed in a 96-well plate at a final concentration of 2 μM in 20 μL volume. Inhibitors were added at a final concentration of 10 μM. SYPRO-orange (Molecular Probes) was added as a fluorescence probe at a dilution of 1 in 5000. Excitation and emission filters were set to 465 and 590 nm, respectively. The temperature was raised with a step of 3 °C per minute, and fluorescence readings were taken at each interval. The temperature dependence of the fluorescence was approximated by the equation

$$Y_{(T)} = y_F + [(y_U + y_F)/(1 + \exp(\Delta uG_{(T)}/RT))]$$

where ΔuG is the difference in unfolding free energy between the folded and unfolded state, R is the gas constant, and y_F and y_U are the fluorescence intensity of the probe in the presence of completely folded and unfolded protein, respectively.[49] The baselines of the denatured and native states were approximated by a linear fit. The observed temperature shifts, ΔT_m^{obs}, for each inhibitor, were recorded as the difference between the transition midpoints of sample and reference wells containing the protein without the inhibitor and were determined by nonlinear least-squares fit. Measurements were performed in triplicate.

4.6. Isothermal Titration Calorimetry. Thermodynamic parameters of the interactions of compounds 30 and 34 with DRAK1 (residues 39—369) were determined by ITC. After equilibration of the sample cell and injection syringe with gel filtration buffer, the compound solution (10—20 μM, prepared in gel filtration buffer) was added into the sample cell and the protein (100 μM in gel filtration buffer) filled into an injection syringe. The experiments were performed at 15 °C with 30 injections of 8 μL each using a Nano ITC (TA Instruments) device.

4.7. NanoBRET Target Engagement Assays. The assay was performed as described previously.[31,50] In brief, full-length kinase ORF cloned in frame with a NanoLuc-vector (Promega, as indicated in Table S10) was transfected into HEK293T cells (ATCC CRL-1573) using FuGENE HD (Promega, E2312), and proteins were allowed to express for 20 h. The serially diluted inhibitor and the NanoBRET kinase tracer (Promega, as indicated in Table S10) were pipetted into white 384-well plates (Greiner 781 207) using an ECHO 550 acoustic dispenser (Labcyte). The corresponding transfected cells were added and reseeded at a density of 2.5 × 10^5 cells/mL after trypsinization and resuspension in Opti-MEM without phenol red (Life Technologies). The system was allowed to equilibrate for 2 h at 37 °C and 5% CO$_2$ prior to BRET measurements. To measure BRET, NanoBRET NanoGlo substrate + extracellular NanoLuc inhibitor (Promega, N2160) were added as per the manufacturer's protocol, and filtered luminescence was measured on a PHERAstar plate reader (BMG Labtech) equipped with a luminescence filter pair [450 nm BP filter (donor) and 610 nm LP filter (acceptor)]. Competitive displacement data were then plotted using GraphPad Prism 9 software using a normalized 3-parameter curve fit with the following equation

$$Y = 100/(1 + 10((X - \log IC_{50})))$$

185

4.8. Cytotoxicity. To assess cytotoxicity as well as potential tubulin binding, a high content screen was performed in three different cell lines using a confocal microscope (CQ1, Yokogawa). HEK293T (ATCC CRL-1573) and U2OS (ATCCHTB-96) were cultured in DMEM plus L-glutamine (high glucose) supplemented by 10% FBS (Gibco) and Penicillin/Streptomycin (Gibco). MRC-9 fibroblasts (ATCC CCL-2) were cultured in EMEM plus L-glutamine supplemented by 10% FBS (Gibco) and Penicillin/Streptomycin (Gibco). For every cell line, 2000 cells per well were seeded in 384-well plates in culture medium (cell culture microplate, PS, f-bottom, μClear, 781091, Greiner). Cells were stained with 60 nM Hoechst33342 (Thermo Scientific), 75 nM Mitotracker red (Invitrogen), 0.3 μL/well Annexin V Alexa Fluor 680 conjugate (Invitrogen) and 25 nL/well BioTracker 488 green microtubule cytoskeleton dye (EMD Millipore). Each compound [CK156 (34) and CK228 (30)] was tested at two different concentrations (1 and 10 μM), $n = 4$. Staurosporine 10 μM was used as a positive control. Cellular shape and fluorescence were measured before compound treatment and 18 h as well as 36 h after compound exposure [CK156 (34), CK228 (30)], respectively, using a CQ1 high-content confocal microscope (Yokogawa). The following setup parameters were used for image acquisition: Ex 405 nm/Em 447/60 nm, 500 ms, 50%; Ex 561 nm/Em 617/73 nm, 100 ms, 40%; Ex 488/Em 525/50 nm, 50 ms, 40%; bright field, 300 ms, 100% transmission, one centered field per well, 7 z-stacks per well with 55 μm spacing. Images were analyzed using the CellPathfinder software (Yokogawa). The cells were detected as described previously,[51] gated using a machine learning algorithm and divided in the categories healthy, apoptotic, and dead cells. The program was trained using a training set of five compounds with known effects (staurosporine, paclitaxel, digitonin, milcidib, and daunorubicin). Data was normalized against the average of DMSO (0.1%) treated cells. Significance was shown using a two-way ANOVA analysis. Two biological duplicates were performed.

4.9. Glioma Cell Viability Assays. The LN-18, LN-229, T98G, and D247-MG human long-term glioma cell lines[52] were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum and 10 μL/mL glutamine (Invitrogen, Basel, Switzerland). For each cell line, 200 cells were seeded in 96-well plates in the culture medium. Compounds CK156, CK228, or CKJB52 were added the following day with concentrations ranging from 0 to 25 μM, $n = 6$. Staurosporine 10 μM was used as a positive control. Cellular metabolism was measured after 7 days of compound exposure using the resazurin reduction method.[53] Resazurin solution (Sigma-Aldrich, 0.15 mg/mL in PBS) was directly added 1:10 to the cells. After 3 h of incubation at 37 °C, resorufin fluorescence was detected using an Infinite 200 Pro (Tecan) plate reader (560 nm excitation/590 nm emission filter set). EC_{50} values were calculated from dose–response curves fitted using a 4-parameter log-logistic model (R package "drc").[54] Cells were imaged on days 1 and 5 following compound exposure using an Axiovert 100 inverted microscope equipped with an AxioCam ICm 1 camera (Zeiss). RNA-seq data downloaded from Cell Model Passports[55] was used to assess DRAK1/STK17A gene expression in these cell lines.

4.10. Organic Synthesis. All commercial chemicals and reagents were obtained from Fluorochem, Merck, TCI, abcr, or Apollo Scientific. Unless otherwise indicated, the purity was ≥95%. The solvents used in the analytical grade were obtained from Fisher Scientific, Merck, and VWR Chemicals and all dry solvents, with AcroSeal septum, from Acros Organics. All performed thin-layer chromatography (TLC) was done with silica gel on aluminum foils (60 Å pore diameter) obtained from Macherey-Nagel and visualized with ultraviolet light ($\lambda = 254$ and 365 nm). The purification via flash chromatography was performed with a Interchim (Montluçon, F) puriFlash XS 420 + equipped with a multi-wavelength UV-DAD (200–400 nm) detector using Interchim PF-30SIHP normal phase columns and PF-30C18HP reversed phase (C_{18}) columns. The nuclear magnetic resonance (NMR) spectroscopy was performed with spectrometers from Bruker (Karlsruhe, Germany) with 126, 151, 250, 400, 500, and 600 MHz. The spectra were recorded in deuterated dimethyl sulfoxide (DMSO-d_6) and the solvent signal was used as a

reference (2.50 ppm). Coupling constants (J) were reported in hertz (Hz) and multiplicities were designated as followed. s (singlet), bs (broad singlet), d (doublet), dd (double doublet), ddd (doublet of doublet of doublets), t (triplet), dt (doublet of triplets), q (quartet), and m (multiplet). Mass spectrometric analysis was carried out with an electrospray ionization device (VG Platform II) from Fisons Instruments (Glasgow, UK). For MALDI-HRMS, a MALDI LTQ Orbitrap XL system from Thermo Scientific (Waltham, MA, USA) was used. Purity of the synthesized compounds was determined with: (I) Agilent Technologies (Santa Clara, CA, USA) 1260 Infinity II device with a 1260 MWD detector (G7165A; 254 nm, 280 nm) and a LC/MSD device (G6125B, ESI pos. 100–1000) using an Agilent Technologies Eclipse XDB-C18 5 μm 4.6 × 250 mm reversed phase (C_{18}) column. As mobile phase Milli-Q water (A) and acetonitrile (B) + 0.1% TFA were used with a flowrate of 1 mL/min. The gradient was running over 19 min starting with 98% A and 2% B, going down on 2% A and 98% B, and finishing at 98% A and 2% B. (II) Agilent Technologies (Santa Clara, CA, USA) 1260 Infinity II device with a 1260 DAD HS detector (G7117C; 254, 280, and 210 nm) and a LC/MSD device (G6125B, ESI pos. 100–1000) using an Agilent Technologies Poroshell 12 EC-C18 2.7 μm 3 × 150 mm reversed phase (C_{18}) column. As mobile phase Milli-Q water (A) and acetonitrile (B) + 0.1% TFA were used with a flowrate of 0.5 mL/min. The gradient was running over 8 min starting with 98% A and 2% B, going down on 2% A and 98% B, and finishing at 98% A and 2% B. The purity of all synthesized final compounds was 95% or higher.

4.10.1. CKJB68 Synthesis of ($1^3Z,1^4E$)-N-Benzyl-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2^4-carboxamide (14). ($1^3Z,1^4E$)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2^4-carboxylic acid (35 mg, 0.09 mmol) was dissolved in dry dimethylformamide (5 mL) and dry N,N-diisopropylethylamine (29 mg, 0.22 mmol) as well as (1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate (41 mg, 0.11 mmol)) were added. The mixture was stirred for 1 h at room temperature under an argon atmosphere. Afterward, benzylamine (12 mg, 0.11 mmol) was added and the mixture was stirred at room temperature for 18 h. After quenching with water and acidification with 1 N HCl, the precipitate was separated by vacuum filtration. The desired product was obtained as a white solid (33 mg, 75%). ^{1}H NMR (500 MHz, DMSO-d_6): δ 8.82 (d, $J = 1.5$ Hz, 1H), 8.64 (t, $J = 6.1$ Hz, 1H), 8.57 (d, $J = 7.6$ Hz, 1H), 8.38 (s, 1H), 7.91 (t, $J = 5.4$ Hz, 1H), 7.83 (d, $J = 8.1$ Hz, 1H), 7.33 (d, $J = 4.5$ Hz, 5H), 7.27–7.19 (m, 1H), 6.34 (d, $J = 7.6$ Hz, 1H), 4.52 (d, $J = 6.1$ Hz, 2H), 4.50–4.43 (m, 2H), 4.04 (dd, $J = 9.1$, 6.3 Hz, 2H), 3.89 (dd, $J = 9.2$, 6.0 Hz, 2H), 3.54 (q, $J = 7.2$, 6.8 Hz, 2H) (ppm). ^{13}C NMR (126 MHz, DMSO-d_6): δ 164.69, 156.50, 156.10, 145.20, 141.66, 140.00, 137.64, 135.80, 131.11, 128.26, 127.02, 126.58, 117.92, 116.20, 110.02, 103.57, 100.15, 65.46, 65.36, 64.03 (ppm). MS (ESI$^+$) m/z: 430.10 [M + H]$^+$. HRMS m/z: [M + H]$^+$ calcd for $C_{24}H_{23}N_5O_3$, 430.18737; found, 430.18729. HPLC (I): t_R = 13.843, purity ≥95%.

4.10.2. CKJB52 Synthesis of ($1^3Z,1^4E$)-N-(2,4-Dichlorobenzyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2^4-carboxamide (15). The title compound was synthesized according to the procedure of 14 using 2,4-dichlorobenzylamine (28 mg, 0.16 mmol). The desired product was obtained as a white solid (53 mg, 72%). ^{1}H NMR (500 MHz, DMSO-d_6): δ 8.84 (d, $J = 1.4$ Hz, 1H), 8.74 (t, $J = 6.1$ Hz, 1H), 8.57 (d, $J = 7.6$ Hz, 1H), 8.39 (s, 1H), 7.92 (t, $J = 5.4$ Hz, 1H), 7.82 (d, $J = 8.1$ Hz, 1H), 7.63 (d, $J = 2.2$ Hz, 1H), 7.44 (dd, $J = 8.4$, 2.2 Hz, 1H), 7.36 (d, $J = 8.4$ Hz, 1H), 7.33 (dd, $J = 8.2$, 1.4 Hz, 1H), 6.34 (d, $J = 7.6$ Hz, 1H), 4.54 (d, $J = 6.1$ Hz, 2H), 4.52–4.47 (m, 2H), 4.12–3.99 (m, 2H), 3.94–3.84 (m, 2H), 3.61–3.48 (m, 2H) (ppm). ^{13}C NMR (126 MHz, DMSO-d_6): δ 164.86, 156.63, 156.13, 145.25, 141.70, 137.92, 136.07, 135.82, 132.68, 131.95, 131.11, 129.84, 128.51, 127.38, 117.46, 116.26, 110.08, 103.55, 100.17, 65.55, 65.40, 64.09 (ppm). MS (ESI$^+$) m/z: 500.10 [M + H]$^+$. HRMS m/z: [M + H]$^+$ calcd for $C_{24}H_{23}Cl_2N_5O_3$, 499.11278; found, 499.11153. HPLC (I): t_R = 16.058, purity ≥95%.

186

4.10.3. CKJB53 Synthesis of (1³Z,1⁴E)-N-(4-Chloro-2-fluorobenzyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacydononaphane-2⁴-carboxamide (16). The title compound was synthesized according to the procedure of 14 using (4-chloro-2-fluorophenyl)methanamine (25 mg, 0.16 mmol). The desired product was obtained as a white solid (53 mg, 75%). ¹H NMR (500 MHz, DMSO-d_6): δ 8.83 (d, J = 1.5 Hz, 1H), 8.68 (t, J = 6.1 Hz, 1H), 8.57 (d, J = 7.6 Hz, 1H), 8.38 (s, 1H), 7.92 (t, J = 5.4 Hz, 1H), 7.81 (d, J = 8.1 Hz, 1H), 7.41 (dd, J = 10.1, 2.1 Hz, 1H), 7.37 (t, J = 8.3 Hz, 1H), 7.32 (dd, J = 8.1, 1.4 Hz, 1H), 7.28 (dd, J = 8.3, 2.0 Hz, 1H), 6.34 (d, J = 7.6 Hz, 1H), 4.55−4.44 (m, 4H), 4.12−3.98 (m, 2H), 3.96−3.83 (m, 2H), 3.60−3.49 (m, 2H) (ppm). ¹³C NMR (126 MHz, DMSO-d_6): δ 164.89, 160.75, 158.78, 156.59, 156.12, 145.24, 141.69, 137.86, 135.81, 131.94, 131.09, 130.39, 126.03, 124.62, 117.56, 116.22, 115.74, 115.54, 110.04, 103.55, 100.17, 65.49, 65.38, 64.05 (ppm). MS (ESI⁺) m/z: 482.10 [M + H]⁺. HRMS m/z: [M + H]⁺ calcd for $C_{24}H_{21}ClFN_5O_3$, 482.13897; found, 482.13833. HPLC (I): t_R = 15.182, purity ≥95%.

4.10.4. CKJB51 Synthesis of (1³Z,1⁴E)-N-(4-Chloro-2-methylbenzyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacydononaphane-2⁴-carboxamide (17). The title compound was synthesized according to the procedure of 14 using (4-chloro-2-methylphenyl)methanamine (25 mg, 0.16 mmol). The desired product was obtained as a white solid (57 mg, 81%). ¹H NMR (500 MHz, DMSO-d_6): δ 8.82 (d, J = 1.5 Hz, 1H), 8.57 (dd, J = 6.8, 5.0 Hz, 1H), 8.38 (s, 1H), 7.91 (t, J = 5.4 Hz, 1H), 7.80 (d, J = 8.1 Hz, 1H), 7.32 (dd, J = 8.1, 1.4 Hz, 1H), 7.28−7.20 (m, 3H), 6.34 (d, J = 7.5 Hz, 1H), 4.52−4.42 (m, 4H), 4.04 (dd, J = 9.2, 6.1 Hz, 2H), 3.89 (dd, J = 9.1, 6.0 Hz, 2H), 3.54 (q, J = 6.8 Hz, 2H), 2.33 (s, 3H) (ppm). ¹³C NMR (126 MHz, DMSO-d_6): δ 164.74, 156.47, 156.09, 145.22, 145.20, 141.67, 137.75, 137.66, 136.71, 135.81, 131.02, 130.80, 129.33, 128.62, 125.58, 117.96, 116.22, 110.05, 103.58, 100.16, 65.49, 65.38, 64.05, 18.41 (ppm). MS (ESI⁺) m/z: 478.10 [M + H]⁺. HRMS: [M + H]⁺ calcd for $C_{25}H_{24}ClN_5O_3$, 478.16404; found, 478.16341. HPLC (I): t_R = 15.436, purity ≥95%.

4.10.5. CK208 Synthesis of (1³Z,1⁴E)-N-(3,5-Dichlorobenzyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2⁴-carboxamide (18). The title compound was synthesized according to the procedure of 14 using (3,5-dichlorophenyl)methanamine (22 mg, 0.12 mmol). The desired product was obtained as a white solid (39 mg, 76%). ¹H NMR (500 MHz, DMSO-d_6): δ 8.83 (d, J = 1.5 Hz, 1H), 8.73 (t, J = 6.2 Hz, 1H), 8.56 (d, J = 7.6 Hz, 1H), 8.37 (s, 1H), 7.90 (t, J = 5.4 Hz, 1H), 7.77 (d, J = 8.0 Hz, 1H), 7.47 (t, J = 2.0 Hz, 1H), 7.39 (d, J = 1.9 Hz, 2H), 7.32 (dd, J = 8.1, 1.4 Hz, 1H), 6.34 (d, J = 7.6 Hz, 1H), 4.56−4.43 (m, 4H), 4.10−4.00 (m, 2H), 3.95−3.83 (m, 2H), 3.60−3.49 (m, 2H) (ppm). ¹³C NMR (126 MHz, DMSO-d_6): δ 165.06, 156.41, 156.08, 145.20, 144.18, 141.62, 137.71, 135.76, 133.83, 130.89, 126.16, 125.85, 117.93, 116.18, 110.00, 103.54, 100.13, 65.43, 65.40, 64.09, 41.80 (ppm). MS (ESI⁺) m/z: 498.20 [M + H]⁺. HRMS m/z: [M + H]⁺ calcd for $C_{24}H_{21}Cl_2N_5O_3$, 498.10942; found, 498.10905. HPLC (I): t_R = 14.624, purity ≥95%.

4.10.6. CK200 Synthesis of (1³Z,1⁴E)-N-(3-Chloro-5-fluorobenzyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacydononaphane-2⁴-carboxamide (19). The title compound was synthesized according to the procedure of 14 using (3-chloro-5-fluorophenyl)methanamine (23 mg, 0.14 mmol). The desired product was obtained as a white solid (30 mg, 53%). ¹H NMR (500 MHz, DMSO-d_6): δ 8.83 (d, J = 1.5 Hz, 1H), 8.74 (t, J = 6.2 Hz, 1H), 8.55 (d, J = 7.5 Hz, 1H), 8.37 (s, 1H), 7.90 (t, J = 5.4 Hz, 1H), 7.78 (d, J = 8.1 Hz, 1H), 7.32 (dd, J = 8.2, 1.4 Hz, 1H), 7.28 (dt, J = 8.6, 2.2 Hz, 1H), 7.26−7.25 (m, 1H), 7.18−7.13 (m, 1H), 6.34 (d, J = 7.6 Hz, 1H), 4.52−4.47 (m, 4H), 4.09−4.00 (m, 2H), 3.93−3.84 (m, 2H), 3.56−3.51 (m, 2H) (ppm). ¹³C NMR (126 MHz, DMSO-d_6): δ 165.55, 163.59, 161.63, 156.94, 156.57, 145.68, 145.47, 145.40, 142.09, 138.22, 136.22, 134.20, 134.11, 131.43, 123.65, 123.63, 118.32, 116.67, 114.63, 114.43, 113.39, 113.21, 110.50, 104.02, 100.63, 65.95, 65.90, 64.59, 42.42 (ppm). MS (ESI⁺) m/z: 482.20 [M + H]⁺. HRMS m/z: [M + H]⁺ calcd for $C_{24}H_{22}ClFN_5O_3$ 482.13897; found, 482.13835. HPLC (I): t_R = 14.031, purity ≥95%.

4.10.7. CK199 Synthesis of (1³Z,1⁴E)-N-(2,3-Dichlorobenzyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2⁴-carboxamide (20). The title compound was synthesized according to the procedure of 14 using (2,3-dichlorophenyl)methanamine (25 mg, 0.14 mmol). The desired product was obtained as a white solid (46 mg, 79%). ¹H NMR (500 MHz, DMSO-d_6): δ 8.85 (d, J = 1.5 Hz, 1H), 8.77 (t, J = 6.1 Hz, 1H), 8.57 (d, J = 7.6 Hz, 1H), 8.38 (s, 1H), 7.91 (t, J = 5.4 Hz, 1H), 7.83 (d, J = 8.1 Hz, 1H), 7.55 (dd, J = 7.8, 1.8 Hz, 1H), 7.40−7.30 (m, 3H), 6.34 (d, J = 7.6 Hz, 1H), 4.60 (d, J = 6.1 Hz, 2H), 4.53−4.48 (m, 2H), 4.10−4.01 (m, 2H), 3.93−3.86 (m, 2H), 3.58−3.51 (m, 2H) (ppm). ¹³C NMR (126 MHz, DMSO-d_6): δ 164.84, 156.63, 156.10, 145.22, 141.65, 139.56, 137.91, 135.77, 131.57, 131.08, 129.66, 128.75, 128.09, 126.94, 117.44, 116.24, 110.09, 103.52, 100.15, 65.61, 65.44, 64.17, 41.39 (ppm). MS (ESI⁺) m/z: 498.10 [M + H]⁺. HRMS m/z: [M + H]⁺ calcd for $C_{24}H_{22}Cl_2N_5O_3$, 498.10942; found, 498.10862. HPLC (I): t_R = 9.533, purity ≥95%.

4.10.8. CK203 Synthesis of (1³Z,1⁴E)-N-(3-Chloro-2-fluorobenzyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2⁴-carboxamide (21). The title compound was synthesized according to the procedure of 14 using (3-chloro-2-fluorophenyl)methanamine (23 mg, 0.14 mmol). The desired product was obtained as a slightly yellow solid (38 mg, 67%). ¹H NMR (500 MHz, DMSO-d_6): δ 8.83 (d, J = 1.5 Hz, 1H), 8.71 (t, J = 6.0 Hz, 1H), 8.55 (d, J = 7.5 Hz, 1H), 8.37 (s, 1H), 7.90 (t, J = 4.9 Hz, 1H), 7.81 (d, J = 8.1 Hz, 1H), 7.46 (td, J = 7.6, 1.7 Hz, 1H), 7.36−7.29 (m, 2H), 7.23−7.17 (m, 1H), 6.34 (d, J = 7.6 Hz, 1H), 4.57 (d, J = 6.0 Hz, 2H), 4.52−4.46 (m, 2H), 4.07−4.01 (m, 2H), 3.88 (dd, J = 9.0, 6.1 Hz, 2H), 3.53 (s, 2H) (ppm). ¹³C NMR (126 MHz, DMSO-d_6): δ 165.38, 157.07, 156.58, 145.69, 142.10, 138.33, 136.22, 131.55, 129.33, 129.27, 129.21, 128.38, 128.35, 125.70, 125.65, 125.62, 117.98, 116.69, 110.55, 104.01, 100.64, 66.05, 64.61, 37.05 (ppm). MS (ESI⁺) m/z: 482.20 [M + H]⁺. HRMS m/z: [M + H]⁺ calcd for $C_{24}H_{22}ClFN_5O_3$, 482.13897; found, 482.13837. HPLC (I): t_R = 13.927, purity ≥95%.

4.10.9. CK209 Synthesis of (1³Z,1⁴E)-N-(Naphthalen-1-ylmethyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2⁴-carboxamide (22). The title compound was synthesized according to the procedure of 14 using naphthalen-1-ylmethanamine (19 mg, 0.12 mmol). The desired product was obtained as a white solid (31 mg, 63%). ¹H NMR (500 MHz, DMSO-d_6): δ 8.82 (d, J = 1.5 Hz, 1H), 8.68 (t, J = 6.0 Hz, 1H), 8.56 (d, J = 7.6 Hz, 1H), 8.37 (s, 1H), 8.22 (dd, J = 8.4, 1.1 Hz, 1H), 7.99−7.94 (m, 1H), 7.91 (t, J = 5.4 Hz, 1H), 7.86−7.81 (m, 2H), 7.62−7.53 (m, 2H), 7.52−7.46 (m, 2H), 7.33 (dd, J = 8.1, 1.4 Hz, 1H), 6.34 (d, J = 7.6 Hz, 1H), 4.98 (d, J = 5.9 Hz, 2H), 4.50−4.36 (m, 2H), 4.07−3.96 (m, 2H), 3.93−3.81 (m, 2H), 3.62−3.45 (m, 2H) (ppm). ¹³C NMR (126 MHz, DMSO-d_6): δ 164.64, 156.46, 156.08, 145.18, 141.61, 137.62, 135.75, 134.94, 133.29, 131.00, 130.75, 128.54, 127.27, 126.15, 125.74, 125.48, 124.77, 123.40, 118.03, 116.21, 110.10, 103.54, 100.12, 65.57, 65.41, 64.12, 40.68 (ppm). MS (ESI⁺) m/z: 480.15 [M + H]⁺. HRMS m/z: [M + H]⁺ calcd for $C_{28}H_{26}N_5O_3$, 480.20302; found, 480.20349. HPLC (II): t_R = 9.264, purity ≥95%.

4.10.10. CK201 Synthesis of (1³Z,1⁴E)-N-(1-Phenylethyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2⁴-carboxamide (23). The title compound was synthesized according to the procedure of 14 using 1-phenylethanamine (17 mg, 0.14 mmol). The desired product was obtained as a white solid (38 mg, 72%). ¹H NMR (500 MHz, DMSO-d_6): δ 8.83 (d, J = 1.5 Hz, 1H), 8.55 (d, J = 7.5 Hz, 1H), 8.38 (d, J = 8.0 Hz, 1H), 8.37 (s, 1H), 7.90 (t, J = 5.4 Hz, 1H), 7.74 (d, J = 8.1 Hz, 1H), 7.41−7.30 (m, 4H), 7.25−7.21 (m, 1H), 6.34 (d, J = 7.6 Hz, 1H), 5.18−5.10 (m, 1H), 4.52−4.44 (m, 2H), 4.08−4.00 (m, 2H), 3.94−3.85 (m, 2H), 3.57−3.52 (m, 2H), 1.47 (d, J = 7.0 Hz, 3H) (ppm). ¹³C NMR (126 MHz, DMSO-d_6): δ 163.91, 156.46, 156.12, 145.22, 144.81, 141.65, 137.56, 135.80, 130.84, 128.39, 126.69, 125.97, 125.94, 118.36, 116.32, 110.30, 103.59, 100.19, 65.89, 65.56, 64.32, 48.44, 35.83, 22.84 (ppm). MS (ESI⁺) m/z: 444.20 [M + H]⁺. HRMS m/z: [M + H]⁺ calcd for $C_{25}H_{26}N_5O_3$, 444.20302; found, 444.20375. HPLC (I): t_R = 13.596, purity ≥95%.

187

4.10.11. CK224 Synthesis of (1³Z,1⁴E)-N-((S)-1-Phenylethyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2⁴-carboxamide (24). The title compound was synthesized according to the procedure of 14 using (S)-1-phenylethanamine (13 mg, 0.11 mmol). The desired product was obtained as a slightly yellow solid (32 mg, 82%). ¹H NMR (500 MHz, DMSO-d_6): δ 8.84 (s, 1H), 8.56 (d, J = 7.5 Hz, 1H), 8.41−8.34 (m, 2H), 7.91 (t, J = 5.4 Hz, 1H), 7.75 (d, J = 8.0 Hz, 1H), 7.41−7.30 (m, 5H), 7.26−7.21 (m, 1H), 6.34 (d, J = 7.6 Hz, 1H), 5.14 (p, J = 7.1 Hz, 1H), 4.56−4.42 (m, 2H), 4.05 (t, J = 7.5 Hz, 2H), 3.95−3.85 (m, 2H), 3.59−3.51 (m, 2H), 1.47 (d, J = 7.0 Hz, 3H) (ppm). ¹³C NMR (126 MHz, DMSO-d_6): δ 164.31, 156.89, 156.55, 145.66, 145.26, 142.09, 137.99, 136.24, 131.27, 128.82, 127.11, 126.40, 118.81, 116.74, 110.72, 104.02, 100.61, 66.28, 65.96, 64.72, 48.86, 23.28 (ppm). MS (ESI⁺) m/z: 444.10 [M + H]⁺. HRMS m/z: [M + Na]⁺ calcd for C₂₅H₂₅N₅NaO₃ 466.18496; found, 466.18413. HPLC (II): t_R = 8.938, purity ≥95%.

4.10.12. CK233 Synthesis of (1³Z,1⁴E)-N-((R)-1-Phenylethyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2⁴-carboxamide (25). The title compound was synthesized according to the procedure of 14 using (R)-1-phenylethanamine (11 mg, 0.09 mmol). The desired product was obtained as a white solid (27 mg, 83%). ¹H NMR (500 MHz, DMSO-d_6): δ 8.84 (d, J = 1.5 Hz, 1H), 8.56 (d, J = 7.6 Hz, 1H), 8.41−8.35 (m, 2H), 7.91 (t, J = 5.4 Hz, 1H), 7.75 (d, J = 8.0 Hz, 1H), 7.42−7.38 (m, 2H), 7.37−7.30 (m, 3H), 7.26−7.21 (m, 1H), 6.34 (d, J = 7.6 Hz, 1H), 5.14 (p, J = 7.1 Hz, 1H), 4.54−4.44 (m, 2H), 4.11−3.99 (m, 2H), 3.95−3.84 (m, 2H), 3.60−3.49 (m, 2H), 1.47 (d, J = 7.0 Hz, 3H) (ppm). ¹³C NMR (126 MHz, DMSO-d_6): δ 163.82, 156.40, 156.06, 145.16, 144.78, 141.60, 137.50, 135.75, 130.78, 128.33, 126.62, 125.92, 118.32, 116.25, 110.23, 103.54, 100.12, 65.80, 65.47, 64.23, 48.37, 22.80 (ppm). MS (ESI⁺) m/z: 444.10 [M + H]⁺. HRMS m/z: [M + H]⁺ calcd for C₂₅H₂₆N₅O₃, 444.20302; found, 444.20273. HPLC (II): t_R = 8.939, purity ≥95%.

4.10.13. CK225 Synthesis of (1³Z,1⁴E)-N-(1-(Naphthalen-1-yl)ethyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2⁴-carboxamide (26). The title compound was synthesized according to the procedure of 14 using 1-(naphthalen-1-yl)ethanamine (18 mg, 0.11 mmol). The desired product was obtained as a white solid (33 mg, 76%). ¹H NMR (500 MHz, DMSO-d_6): δ 8.83 (d, J = 1.5 Hz, 1H), 8.60−8.52 (m, 2H), 8.36 (s, 1H), 8.23 (d, J = 8.4 Hz, 1H), 7.96 (dd, J = 8.2, 1.4 Hz, 1H), 7.90 (t, J = 5.4 Hz, 1H), 7.83 (d, J = 8.1 Hz, 1H), 7.73 (d, J = 8.0 Hz, 1H), 7.65−7.48 (m, 4H), 7.31 (dd, J = 8.1, 1.4 Hz, 1H), 6.33 (d, J = 7.6 Hz, 1H), 5.93 (p, J = 7.1 Hz, 1H), 4.51−4.38 (m, 2H), 4.09−3.96 (m, 2H), 3.93−3.83 (m, 2H), 3.59−3.49 (m, 2H), 1.61 (d, J = 6.9 Hz, 3H) (ppm). ¹³C NMR (126 MHz, DMSO-d_6): δ 163.94, 156.44, 156.10, 145.21, 141.63, 140.50, 137.53, 135.78, 133.48, 130.75, 130.23, 128.75, 127.28, 126.19, 125.64, 123.22, 122.59, 118.49, 116.30, 110.29, 103.59, 100.17, 65.84, 65.53, 64.28, 45.03, 22.12 (ppm). MS (ESI⁺) m/z: 494.05 [M + H]⁺. HRMS m/z: [M + Na]⁺ calcd for C₂₉H₂₇N₅NaO₃, 516.20061; found, 516.20013. HPLC (II): t_R = 9.589, purity ≥95%.

4.10.14. CK207 Synthesis of (1³Z,1⁴E)-N-(Pyridin-3-ylmethyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2⁴-carboxamide (27). The title compound was synthesized according to the procedure of 14 using pyridin-3-ylmethanamine (15 mg, 0.14 mmol). The desired product was obtained as a white solid (40 mg, 79%). ¹H NMR (500 MHz, DMSO-d_6): δ 8.82 (d, J = 1.5 Hz, 1H), 8.73 (t, J = 6.1 Hz, 1H), 8.55 (d, J = 7.5 Hz, 2H), 8.47−8.42 (m, 1H), 8.37 (s, 1H), 7.90 (t, J = 5.4 Hz, 1H), 7.82 (d, J = 8.1 Hz, 1H), 7.73 (dt, J = 7.9, 2.0 Hz, 1H), 7.36 (dd, J = 7.9, 4.8 Hz, 1H), 7.31 (dd, J = 8.1, 1.4 Hz, 1H), 6.33 (d, J = 7.6 Hz, 1H), 4.53 (d, J = 6.1 Hz, 2H), 4.50−4.44 (m, 2H), 4.08−3.99 (m, 2H), 3.92−3.83 (m, 2H), 3.53 (q, J = 6.7 Hz, 2H) (ppm). ¹³C NMR (126 MHz, DMSO-d_6): δ 164.91, 156.56, 156.11, 148.70, 147.85, 145.23, 141.66, 137.77, 135.78, 135.57, 135.02, 131.10, 123.50, 117.72, 116.20, 110.04, 103.57, 100.17, 65.51, 65.44, 64.11, 40.44 (ppm). MS (ESI⁺) m/z: 431.20 [M + H]⁺. HRMS m/z: [M + H]⁺ calcd for C₂₃H₂₃N₆O₃, 431.18262; found, 431.18114. HPLC (I): t_R = 11.177, purity ≥95%.

4.10.15. CK205 Synthesis of (1³Z,1⁴E)-N-(Pyridin-4-ylmethyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2⁴-carboxamide (28). The title compound was synthesized according to the procedure of 14 using pyridin-4-ylmethanamine (15 mg, 0.14 mmol). The desired product was obtained as a white solid (42 mg, 83%). ¹H NMR (500 MHz, DMSO-d_6): δ 8.84 (d, J = 1.5 Hz, 1H), 8.74 (t, J = 6.1 Hz, 1H), 8.56 (d, J = 7.6 Hz, 1H), 8.53−8.47 (m, 2H), 8.38 (s, 1H), 7.91 (t, J = 5.4 Hz, 1H), 7.82 (d, J = 8.1 Hz, 1H), 7.35−7.28 (m, 3H), 6.34 (d, J = 7.6 Hz, 1H), 4.57−4.46 (m, 4H), 4.09−4.02 (m, 2H), 3.93−3.86 (m, 2H), 3.59−3.50 (m, 2H) (ppm). ¹³C NMR (126 MHz, DMSO-d_6): δ 164.96, 156.57, 156.09, 149.43, 149.07, 145.20, 141.63, 137.78, 135.76, 131.08, 122.00, 117.64, 116.17, 110.02, 103.54, 100.14, 65.48, 65.40, 64.08, 41.83 (ppm). MS (ESI⁺) m/z: 431.20 [M + H]⁺. HRMS m/z: [M + H]⁺ calcd for C₂₃H₂₃N₆O₃, 431.18262; found, 431.18294. HPLC (I): t_R = 11.156, purity ≥95%.

4.10.16. CK227 Synthesis of (1³Z,1⁴E)-N-((R)-1-(Pyridin-4-yl)ethyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2⁴-carboxamide (29). The title compound was synthesized according to the procedure of 14 using (R)-1-(pyridin-4-yl)ethanamine (13 mg, 0.11 mmol). The desired product was obtained as a dark yellow solid (34 mg, 87%). ¹H NMR (500 MHz, DMSO-d_6): δ 8.84 (d, J = 1.5 Hz, 1H), 8.56 (d, J = 7.5 Hz, 1H), 8.54−8.50 (m, 2H), 8.44 (d, J = 7.6 Hz, 1H), 8.37 (s, 1H), 7.92 (t, J = 5.4 Hz, 1H), 7.71 (d, J = 8.0 Hz, 1H), 7.41−7.36 (m, 2H), 7.31 (dd, J = 8.1, 1.3 Hz, 1H), 6.34 (d, J = 7.6 Hz, 1H), 5.12 (p, J = 7.1 Hz, 1H), 4.56−4.45 (m, 2H), 4.11−4.02 (m, 2H), 3.94−3.85 (m, 2H), 3.59−3.48 (m, 2H), 1.48 (d, J = 7.1 Hz, 3H) (ppm). ¹³C NMR (126 MHz, DMSO-d_6): δ 164.31, 156.40, 156.08, 153.53, 149.60, 145.18, 141.60, 135.75, 130.69, 121.24, 118.30, 116.20, 110.18, 103.54, 100.13, 65.73, 65.46, 64.20, 47.73, 21.84 (ppm). MS (ESI⁺) m/z: 445.10 [M + H]⁺. HRMS m/z: [M + Na]⁺ calcd for C₂₄H₂₄N₆NaO₃, 467.18021; found, 467.17882. HPLC (II): t_R = 6.782, purity ≥95%.

4.10.17. CK228 Synthesis of (1³Z,1⁴E)-N-((S)-1-(Pyridin-4-yl)ethyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2⁴-carboxamide (30). The title compound was synthesized according to the procedure of 14 using (S)-1-(pyridin-4-yl)ethanamine (13 mg, 0.11 mmol). The desired product was obtained as a slightly yellow solid (32 mg, 82%). ¹H NMR (500 MHz, DMSO-d_6): δ 8.84 (d, J = 1.5 Hz, 1H), 8.56 (d, J = 7.6 Hz, 1H), 8.54−8.50 (m, 2H), 8.44 (d, J = 7.7 Hz, 1H), 8.37 (s, 1H), 7.92 (t, J = 5.4 Hz, 1H), 7.71 (d, J = 8.1 Hz, 1H), 7.41−7.37 (m, 2H), 7.31 (dd, J = 8.0, 1.4 Hz, 1H), 6.34 (d, J = 7.6 Hz, 1H), 5.12 (p, J = 7.1 Hz, 1H), 4.56−4.45 (m, 2H), 4.14−4.02 (m, 2H), 3.95−3.85 (m, 2H), 3.61−3.49 (m, 2H), 1.48 (d, J = 7.0 Hz, 3H) (ppm). ¹³C NMR (126 MHz, DMSO-d_6): δ 164.31, 156.40, 156.08, 153.52, 149.60, 145.18, 141.60, 137.57, 135.75, 130.69, 121.24, 118.30, 116.20, 110.18, 103.54, 100.13, 65.73, 65.46, 64.20, 47.73, 21.84 (ppm). MS (ESI⁺) m/z: 445.15 [M + H]⁺. HRMS m/z: [M + H]⁺ calcd for C₂₄H₂₅N₆O₃, 445.19827; found, 445.19843. HPLC (II): t_R = 6.788, purity ≥95%.

4.10.18. CK226 Synthesis of (1³Z,1⁴E)-N-(Quinolin-4-ylmethyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2⁴-carboxamide (31). The title compound was synthesized according to the procedure of 14 using quinolin-4-ylmethanamine (21 mg, 0.11 mmol). The desired product was obtained as a slightly yellow solid (30 mg, 71%). ¹H NMR (500 MHz, DMSO-d_6): δ 8.89 (d, J = 4.5 Hz, 1H), 8.86−8.81 (m, 2H), 8.57 (d, J = 7.5 Hz, 1H), 8.39 (s, 1H), 8.31 (dd, J = 8.4, 1.4 Hz, 1H), 8.08 (dd, J = 8.4, 1.3 Hz, 1H), 7.91 (t, J = 5.4 Hz, 1H), 7.85−7.81 (m, 2H), 7.74−7.69 (m, 1H), 7.47 (d, J = 4.6 Hz, 1H), 7.34 (dd, J = 8.1, 1.4 Hz, 1H), 6.35 (d, J = 7.6 Hz, 1H), 5.05 (d, J = 5.9 Hz, 2H), 4.56−4.46 (m, 2H), 4.09−4.03 (m, 2H), 3.93−3.87 (m, 2H), 3.59−3.52 (m, 2H) (ppm). ¹³C NMR (126 MHz, DMSO-d_6): δ 165.03, 156.57, 156.10, 145.22, 141.64, 137.81, 135.78, 131.04, 126.84, 125.99, 123.69, 118.84, 117.75, 116.21, 110.08, 103.55, 100.15, 65.53, 65.43, 64.12 (ppm). MS (ESI⁺) m/z: 481.15 [M + H]⁺. HRMS m/z: [M + Na]⁺ calcd for C₂₇H₂₄N₆NaO₃, 503.18021; found, 503.17904. HPLC (II): t_R = 7.067, purity ≥95%.

4.10.19. CK183 Synthesis of (1³Z,1⁴E)-N,N-Dimethyl-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2⁴-carboxamide (32). The title compound was synthesized

according to the procedure of 14 using ammonia 2 M in ethanol (88 μL, 0.18 mmol). The desired product was obtained as a white solid (23 mg, 42%). ^{1}H NMR (500 MHz, DMSO-d_6): δ 8.80 (d, J = 1.5 Hz, 1H), 8.56 (d, J = 7.6 Hz, 1H), 8.37 (s, 1H), 7.90 (t, J = 5.5 Hz, 1H), 7.86 (d, J = 8.1 Hz, 1H), 7.58 (s, 1H), 7.42 (s, 1H), 7.31 (dd, J = 8.1, 1.4 Hz, 1H), 6.33 (d, J = 7.6 Hz, 1H), 4.54–4.40 (m, 2H), 4.07–3.99 (m, 2H), 3.93–3.83 (m, 2H), 3.61–3.48 (m, 2H) (ppm). ^{13}C NMR (126 MHz, DMSO-d_6): δ 165.81, 156.71, 156.07, 145.17, 141.63, 137.82, 135.76, 131.34, 117.70, 116.06, 109.88, 103.55, 100.12, 65.31, 63.95 (ppm). MS (ESI$^+$) m/z: 340.20 [M + H]$^+$. HRMS m/z: [M + H]$^+$ calcd for $C_{17}H_{18}N_5O_3$, 340.14042; found, 340.14063. HPLC (I): t_R = 9.218, purity ≥95%.

4.10.20. CK182 Synthesis of ($1^3Z,1^4E$)-N-Methyl-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2^4-carboxamide (33). The title compound was synthesized according to the procedure of 14 using methylamine 33% in ethanol (22 μL, 0.18 mmol). The desired product was obtained as a white solid (30 mg, 58%). ^{1}H NMR (500 MHz, DMSO-d_6): δ 8.79 (d, J = 1.5 Hz, 1H), 8.56 (d, J = 7.5 Hz, 1H), 8.36 (s, 1H), 8.07 (q, J = 4.6 Hz, 1H), 7.89 (t, J = 5.4 Hz, 1H), 7.83 (d, J = 8.1 Hz, 1H), 7.30 (dd, J = 8.1, 1.5 Hz, 1H), 6.33 (d, J = 7.6 Hz, 1H), 4.51–4.42 (m, 2H), 4.06–4.00 (m, 2H), 3.92–3.85 (m, 2H), 3.58–3.49 (m, 2H), 2.81 (d, J = 4.7 Hz, 3H) (ppm). ^{13}C NMR (126 MHz, DMSO-d_6): δ 164.96, 156.34, 156.04, 145.14, 141.60, 137.41, 135.75, 131.00, 117.94, 116.06, 109.84, 103.55, 100.11, 65.32, 65.25, 63.91, 26.34 (ppm). MS (ESI$^+$) m/z: 354.20 [M + H]$^+$. HRMS m/z: [M + H]$^+$ calcd for $C_{18}H_{19}N_5O_3$, 354.15607; found, 354.15544. HPLC (I): t_R = 9.411, purity ≥95%.

4.10.21. CK156 Synthesis of ($1^3Z,1^4E$)-N-(tert-Butyl)-3,6-dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclonona-phane-2^4-carboxamide (34). The title compound was synthesized according to the procedure of 14 using 2-methylpropan-2-amine (12 mg, 0.16 mmol). The desired product was obtained as a beige solid (39 mg, 67%). ^{1}H NMR (500 MHz, DMSO-d_6): δ 8.82 (d, J = 1.5 Hz, 1H), 8.56 (d, J = 7.6 Hz, 1H), 8.37 (s, 1H), 7.90 (t, J = 5.4 Hz, 1H), 7.86–7.81 (m, 2H), 7.32 (dd, J = 8.1, 1.4 Hz, 1H), 6.33 (d, J = 7.6 Hz, 1H), 4.52–4.41 (m, 2H), 4.08–3.98 (m, 2H), 3.93–3.83 (m, 2H), 3.60–3.49 (m, 2H), 1.38 (s, 9H) (ppm). ^{13}C NMR (126 MHz, DMSO-d_6): δ 163.43, 156.29, 156.05, 145.13, 141.60, 137.44, 135.74, 130.81, 118.65, 116.34, 110.26, 103.48, 100.09, 65.97, 65.48, 64.27, 50.22, 28.67 (ppm). MS (ESI$^+$) m/z: 396.20 [M + H]$^+$. HRMS m/z: [M + H]$^+$ calcd for $C_{21}H_{26}N_5O_3$, 396.20302; found, 396.20288. HPLC (I): t_R = 13.293, purity ≥95%.

4.10.22. CKJB71 Synthesis of (($1^3Z,1^4E$)-3,6-Dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2^4-yl)(4-methylpiperazin-1-yl)methanone (35). The synthesis of the title compound was carried out according to the procedure of 14 using 1-methylpiperazine (16 mg, 1.2 equiv, 0.16 mmol). Purification was done by flash reversed-phase column chromatography with a mobile phase of dichloromethane and methanol (isocratic ratio 93:7). The desired product was obtained as a white solid (55 mg, 89%). ^{1}H NMR (250 MHz, DMSO-d_6): δ 8.72 (d, J = 1.3 Hz, 1H, PhH), 8.54 (d, J = 7.6 Hz, 1H, HetH), 8.33 (s, 1H, HetH), 7.86 (t, J = 5.2 Hz, 1H, NH), 7.25 (dd, J = 7.8, 1.2 Hz, 1H, PhH), 7.13 (d, J = 7.8 Hz, 1H, PhH), 6.32 (d, J = 7.6 Hz, 1H, HetH), 4.42–4.25 (m, 2H, CH$_2$), 4.08–3.77 (m, 4H, CH$_2$), 3.67–3.47 (m, 4H, CH$_2$), 3.23–3.11 (m, 2H, CH$_2$), 2.45–2.17 (m, 7H, CH$_2$, CH$_3$) (ppm). ^{13}C NMR (151 MHz, DMSO-d_6): δ 166.49, 155.83, 153.71, 144.93, 141.27, 135.64, 135.42, 128.13, 121.38, 116.14, 109.61, 103.72, 99.98, 65.27, 64.70, 63.83, 61.98, 55.99, 54.67, 54.19, 46.04, 45.42 (ppm). MS (ESI$^+$) m/z: 423.40 [M + H]$^+$. HRMS m/z: [M + H]$^+$ calcd for $C_{22}H_{26}N_6O_3$, 425.19586; found, 445.19612. HPLC (I): t_R = 8.139, purity ≥95%.

4.10.23. CKJB75 Synthesis of (($1^3Z,1^4E$)-3,6-Dioxa-9-aza-1(3,5)-pyrazolo[1,5-a]pyrimidina-2(1,3)-benzenacyclononaphane-2^4-yl)(morpholino)methanone (36). The synthesis of the title compound was carried out according to the procedure of 14 using morpholine (14 mg, 1.2 equiv, 0.16 mmol). The desired product was obtained as a white solid (54 mg, 90%). ^{1}H NMR (500 MHz, DMSO-d_6): δ 8.73 (d, J = 1.4 Hz, 1H, PhH), 8.55 (d, J = 7.5 Hz, 1H, HetH), 8.33 (s, 1H, HetH), 7.87 (t, J = 5.4 Hz, 1H, NH), 7.26 (dd, J = 7.8, 1.3 Hz, 1H, PhH), 7.17 (d, J = 7.8 Hz, 1H, PhH), 6.32 (d, J = 7.6 Hz, 1H, HetH), 4.42–4.27 (m, 2H, CH$_2$), 4.07–3.79 (m, 4H, CH$_2$), 3.69–3.57 (m,

4H, CH$_2$), 3.57–3.44 (m, 4H, CH$_2$), 3.23–3.11 (m, 2H, CH$_2$) (ppm). ^{13}C NMR (126 MHz, DMSO-d_6): δ 166.62, 155.89, 153.72, 144.98, 141.40, 135.75, 135.60, 128.39, 121.01, 116.25, 109.52, 103.74, 100.04, 66.37, 66.10, 65.11, 64.59, 63.65, 48.61, 46.91, 41.67 (ppm). MS (ESI$^+$) m/z: 410.14 [M + H]$^+$. HRMS m/z [M + H]$^+$ calcd for $C_{21}H_{23}N_5O_4$, 410.18228; found, 410.18262. HPLC (I): t_R = 11.067, purity ≥95%..

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jmedchem.2c00173.

> Synthesis of macrocyclic scaffold 13; DSF assay of compounds 14–36; KINOMEscan of 13, 14, 23, 30, and 34; NanoBRET assay information; ^{33}PanQinase assay information; data collection and refinement statistics; live cell high content screen of 34 and 30; cell viability assay of 30 and 34; and analytical data for compounds 14–36 (PDF)
>
> Molecular formula strings (CSV)
>
> PDB validation report of the STK17 (DRAK1) kinase domain bound to CK156(PDF)
>
> PDB validation report of the STK17 (DRAK1) kinase domain bound to CKjB68 (PDF)

Accession Codes

PDB ID codes: coordinates and structure factors of the DRAK1-inhibitor complexes are available in the Protein Data Bank (PDB) under accession codes 7QUE (DRAK1-14 complex), 7QUF (DRAK1-34 complex). Authors will release the atomic coordinates upon article publication.

■ AUTHOR INFORMATION

Corresponding Authors

Stefan Knapp — *Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Frankfurt 60438, Germany; Structural Genomics Consortium, Buchman Institute for Molecular Life Science (BMLS), Frankfurt 60438, Germany;* orcid.org/0000-0001-5995-6494; Phone: +49 69 798 29871; Email: knapp@pharmchem.uni-frankfurt.de

Thomas Hanke — *Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Frankfurt 60438, Germany; Structural Genomics Consortium, Buchman Institute for Molecular Life Science (BMLS), Frankfurt 60438, Germany;* orcid.org/0000-0001-7202-9468; Email: hanke@pharmchem.uni-frankfurt.de

Authors

Christian G. Kurz — *Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Frankfurt 60438, Germany; Structural Genomics Consortium, Buchman Institute for Molecular Life Science (BMLS), Frankfurt 60438, Germany*

Franziska Preuss — *Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Frankfurt 60438, Germany; Structural Genomics Consortium, Buchman Institute for Molecular Life Science (BMLS), Frankfurt 60438, Germany*

Amelie Tjaden — *Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Frankfurt 60438, Germany; Structural Genomics Consortium, Buchman Institute for Molecular Life Science (BMLS), Frankfurt 60438, Germany*

Martin Cusack — *Department of Neurology and Clinical Neuroscience Center, University Hospital Zurich and University of Zurich, Zurich 8091, Switzerland*

189

Jennifer Alisa Amrhein — Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Frankfurt 60438, Germany; Structural Genomics Consortium, Buchman Institute for Molecular Life Science (BMLS), Frankfurt 60438, Germany

Deep Chatterjee — Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Frankfurt 60438, Germany; Structural Genomics Consortium, Buchman Institute for Molecular Life Science (BMLS), Frankfurt 60438, Germany

Sebastian Mathea — Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Frankfurt 60438, Germany; Structural Genomics Consortium, Buchman Institute for Molecular Life Science (BMLS), Frankfurt 60438, Germany; orcid.org/0000-0001-8500-4569

Lena Marie Berger — Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Frankfurt 60438, Germany; Structural Genomics Consortium, Buchman Institute for Molecular Life Science (BMLS), Frankfurt 60438, Germany; orcid.org/0000-0002-7835-8067

Benedict-Tilman Berger — Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Frankfurt 60438, Germany; Structural Genomics Consortium, Buchman Institute for Molecular Life Science (BMLS), Frankfurt 60438, Germany; orcid.org/0000-0002-3314-2617

Andreas Krämer — Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Frankfurt 60438, Germany; Structural Genomics Consortium, Buchman Institute for Molecular Life Science (BMLS), Frankfurt 60438, Germany; Frankfurt Cancer Institute (FCI), Frankfurt 60596, Germany

Michael Weller — Department of Neurology and Clinical Neuroscience Center, University Hospital Zurich and University of Zurich, Zurich 8091, Switzerland

Tobias Weiss — Department of Neurology and Clinical Neuroscience Center, University Hospital Zurich and University of Zurich, Zurich 8091, Switzerland

Susanne Müller — Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Frankfurt 60438, Germany; Structural Genomics Consortium, Buchman Institute for Molecular Life Science (BMLS), Frankfurt 60438, Germany; orcid.org/0000-0003-2402-4157

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jmedchem.2c00173

Author Contributions

C.G.K., S.M., S.K., and T.H. designed the project; C.G.K. and J.A.A. synthesized the compounds; F.P., D.C., and S.M. expressed, purified, and cocrystallized DRAK1; F.P. and S.M. refined the structures of DRAK1 and the inhibitor complexes; F.P. performed ITC; L.M.B. and B.-T.B. performed Nano-BRET assays; A.K. expressed and purified proteins for the DSF-Panel; A.T. performed cytotoxicity assays; M.C., M.W., and T.W. performed Glioma cell viability assays; S.M., S.K., and T.H. supervised the research. The manuscript was written by C.G.K., S.M., S.K., and T.H. with contributions from all coauthors.

Notes

The authors declare the following competing financial interest(s): B.-T.B. is the founder and CEO of a CRO company named CELLinib GmbH.

■ ABBREVIATIONS

AAK1, AP2-associated protein kinase 1; ANT2, adenine nucleotide translocator 2; BIKE, BMP-2-inducible protein kinase; BMP2K, BMP-2-inducible protein kinase; CK2, casein kinase II; DAPK, death-associated protein kinases; DAPK1, death-associated protein kinase 1; DAPK2, death-associated protein kinase 2; DAPK3, death-associated protein kinase 3; DRAK1, DAP kinase-related apoptosis-inducing protein kinase 1; DRAK2, DAP kinase-related apoptosis-inducing protein kinase 2; GAK, cyclin-G-associated kinase; GBM, glioblastoma multiforme; HNSCC, head and neck squamous cell carcinoma; IKKε, inhibitor of nuclear factor kappa-B kinase subunit epsilon; MNK1, MAP kinase-interacting serine/threonine-protein kinase 1; MNK2, MAP kinase-interacting serine/threonine-protein kinase 2; PKC, protein kinase C; STK17A, serine/threonine kinase 17A; TBK1, serine/threonine-protein kinase TBK1; TGF-β1, transforming growth factor beta

■ REFERENCES

(1) Müller, S.; Chaikuad, A.; Gray, N. S.; Knapp, S. The Ins and Outs of Selective Kinase Inhibitor Development. Nat. Chem. Biol. 2015, 11, 818−821.

(2) Wu, P.; Nielsen, T. E.; Clausen, M. H. Small-Molecule Kinase Inhibitors: An Analysis of FDA-Approved Drugs. Drug Discovery Today 2016, 21, 5−10.

(3) Zhang, J.; Yang, P. L.; Gray, N. S. Targeting Cancer with Small Molecule Kinase Inhibitors. Nat. Rev. Cancer 2009, 9, 28−39.

(4) Ghoreschi, K.; Laurence, A.; O'Shea, J. J. Selectivity and Therapeutic Inhibition of Kinases: To Be or Not to Be? Nat. Immunol. 2009, 10, 356−360.

(5) Attwood, M. M.; Fabbro, D.; Sokolov, A. V.; Knapp, S.; Schiöth, H. B. Trends in Kinase Drug Discovery: Targets, Indications and Inhibitor Design. Nat. Rev. Drug Discovery 2021, 20, 839−861.

(6) Ferguson, F. M.; Gray, N. S. Kinase Inhibitors: The Road Ahead. Nat. Rev. Drug Discovery 2018, 17, 353−377.

(7) Fabbro, D.; Cowan-Jacob, S. W.; Moebitz, H. Ten Things You Should Know about Protein Kinases: IUPHAR Review 14. Br. J. Pharmacol. 2015, 172, 2675−2700.

7815

pubs.acs.org/jmc

(8) Norman, R. A.; Toader, D.; Ferguson, A. D. Structural Approaches to Obtain Kinase Selectivity. *Trends Pharmacol. Sci.* 2012, 33, 273−278.

(9) Marsault, E.; Peterson, M. L. Macrocycles Are Great Cycles: Applications, Opportunities, and Challenges of Synthetic Macrocycles in Drug Discovery. *J. Med. Chem.* 2011, 54, 1961−2004.

(10) Amrhein, J. A.; Knapp, S.; Hanke, T. Synthetic Opportunities and Challenges for Macrocyclic Kinase Inhibitors. *J. Med. Chem.* 2021, 64, 7991−8009.

(11) Driggers, E. M.; Hale, S. P.; Lee, J.; Terrett, N. K. The Exploration of Macrocycles for Drug Discovery - An Underexploited Structural Class. *Nat. Rev. Drug Discovery* 2008, 7, 608−624.

(12) Giordanetto, F.; Kihlberg, J. Macrocyclic Drugs and Clinical Candidates: What Can Medicinal Chemists Learn from Their Properties? *J. Med. Chem.* 2014, 57, 278−295.

(13) Mallinson, J.; Collins, I. Macrocycles in New Drug Discovery. *Future Med. Chem.* 2012, 4, 1409−1438.

(14) Berginski, M. E.; Moret, N.; Liu, C.; Goldfarb, D.; Sorger, P. K.; Gomez, S. M. The Dark Kinase Knowledgebase: An Online Compendium of Knowledge and Experimental Results of Under-studied Kinases. *Nucleic Acids Res.* 2020, 49, D529−D535.

(15) Inbal, B.; Bialik, S.; Sabanay, I.; Shani, G.; Kimchi, A. DAP Kinase and DRP-1 Mediate Membrane Blebbing and the Formation of Autophagic Vesides during Programmed Cell Death. *J. Cell Biol.* 2002, 157, 455−468.

(16) Chen, C.-H.; Wang, W.-J.; Kuo, J.-C.; Tsai, H.-C.; Lin, J.-R.; Chang, Z.-F.; Chen, R.-H. Bidirectional Signals Transduced by DAPK-ERK Interaction Promote the Apoptotic Effect of DAPK. *EMBO J.* 2005, 24, 294−304.

(17) Farag, A. K.; Roh, E. J. Death-Associated Protein Kinase (DAPK) Family Modulators: Current and Future Therapeutic Outcomes. *Med. Res. Rev.* 2019, 39, 349−385.

(18) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* 2011, 7, 539.

(19) Bialik, S.; Kimchi, A. The Death-Associated Protein Kinases: Structure, Function, and Beyond. *Annu. Rev. Biochem.* 2006, 75, 189−210.

(20) Sanjo, H.; Kawai, T.; Akira, S. DRAKs, Novel Serine/Threonine Kinases Related to Death-Associated Protein Kinase That Trigger Apoptosis. *J. Biol. Chem.* 1998, 273, 29066−29071.

(21) Mao, P.; Hever, M. P.; Niemaszyk, L. M.; Haghkerdar, J. M.; Yanco, E. G.; Desai, D.; Beyrouthy, M. J.; Kerley-Hamilton, J. S.; Freemantle, S. J.; Spinella, M. J. Serine/Threonine Kinase 17A Is a Novel P53 Target Gene and Modulator of Cisplatin Toxicity and Reactive Oxygen Species in Testicular Cancer Cells. *J. Biol. Chem.* 2011, 286, 19381−19391.

(22) Oue, Y.; Murakami, S.; Isshiki, K.; Tsuji, A.; Yuasa, K. Intracellular Localization and Binding Partners of Death Associated Protein Kinase-Related Apoptosis-Inducing Protein Kinase 1. *Biochem. Biophys. Res. Commun.* 2018, 496, 1222−1228.

(23) Park, Y.; Kim, W.; Lee, J.-M.; Park, J.; Cho, J. K.; Pang, K.; Lee, J.; Kim, D.; Park, S.-W.; Yang, K.-M.; Kim, S.-J. Cytoplasmic DRAK1 Overexpressed in Head and Neck Cancers Inhibits TGF-B1 Tumor Suppressor Activity by Binding to Smad3 to Interrupt Its Complex Formation with Smad4. *Oncogene* 2015, 34, 5037−5045.

(24) Chen, A. S.; Wardwell-Ozgo, J.; Shah, N. N.; Wright, D.; Appin, C. L.; Vigneswaran, K.; Brat, D. J.; Komblum, H. I.; Read, R. D. DRAK/STK17a Drives Neoplastic Glial Proliferation through Modulation of MRLC Signaling. *Cancer Res.* 2019, 79, 1085−1097.

(25) Mao, P.; Hever-Jardine, M. P.; Rahme, G. J.; Yang, E.; Tam, J.; Kodali, A.; Biswal, B.; Fadul, C. E.; Gaur, A.; Israel, M. A.; Spinella, M. J. Serine/Threonine Kinase 17A Is a Novel Candidate for Therapeutic Targeting in Glioblastoma. *PLoS One* 2013, 8, No. e81803.

(26) Reich, S. H.; Sprengeler, P. A.; Chiang, G. G.; Appleman, J. R.; Chen, J.; Clarine, J.; Eam, B.; Ernst, J. T.; Han, Q.; Goel, V. K.; Han, E. Z. R.; Huang, V.; Hung, I. N. J.; Jemison, A.; Jessen, K. A.; Molter, J.; Murphy, D.; Neal, M.; Parker, G. S.; Shaghafi, M.; Sperry, S.; Staunton, J.; Stumpf, C. R.; Thompson, P. A.; Tran, C.; Webber, S. E.; Wegerski, C. J.; Zheng, H.; Webster, K. R. Structure-Based Design of Pyridone-Aminal EFT508 Targeting Dysregulated Translation by Selective Mitogen-Activated Protein Kinase Interacting Kinases 1 and 2 (MNK1/2) Inhibition. *J. Med. Chem.* 2018, 61, 3516−3540.

(27) Howard, S.; Berdini, V.; Boulstridge, J. A.; Carr, M. G.; Cross, D. M.; Curry, J.; Devine, L. A.; Early, T. R.; Fazal, L.; Gill, A. L.; Heathcote, M.; Maman, S.; Matthews, J. E.; McMenamin, R. L.; Navarro, E. F.; O'Brien, M. A.; O'Reilly, M.; Rees, D. C.; Reule, M.; Tisi, D.; Williams, G.; Vinković, M.; Wyatt, P. G. Fragment-Based Discovery of the Pyrazol-4-Yl Urea (AT9283), a Multitargeted Kinase Inhibitor with Potent Aurora Kinase Activity. *J. Med. Chem.* 2009, 52, 379−388.

(28) Lefranc, J.; Schulze, V. K.; Hillig, R. C.; Briem, H.; Prinz, F.; Mengel, A.; Heinrich, T.; Balint, J.; Rengachari, S.; Irlbacher, H.; Stöckigt, D.; Börner, U.; Bader, B.; Gradl, S. N.; Nising, C. F.; Von Nussbaum, F.; Mumberg, D.; Panne, D.; Wengner, A. M. Discovery of BAY-985, a Highly Selective TBK1/IKKε Inhibitor. *J. Med. Chem.* 2020, 63, 601−612.

(29) Picado, A.; Chaikuad, A.; Wells, C. I.; Shrestha, S.; Zuercher, W. J.; Pickett, J. E.; Kwarcinski, F. E.; Sinha, P.; de Silva, C. S.; Zutshi, R.; Liu, S.; Kannan, N.; Knapp, S.; Drewry, D. H.; Willson, T. M. A Chemical Probe for Dark Kinase STK17B Derives Its Potency and High Selectivity through a Unique P-Loop Conformation. *J. Med. Chem.* 2020, 63, 14626−14646.

(30) Bullock, A. N.; Debreczeni, J. É.; Fedorov, O. Y.; Nelson, A.; Marsden, B. D.; Knapp, S. Structural Basis of Inhibitor Specificity of the Human Protooncogene Proviral Insertion Site in Moloney Murine Leukemia Virus (PIM-1) Kinase. *J. Med. Chem.* 2005, 48, 7604−7614.

(31) Krämer, A.; Kurz, C. G.; Berger, B. T.; Celik, I. E.; Tjaden, A.; Greco, F. A.; Knapp, S.; Hanke, T. Optimization of Pyrazolo[1,5-a]Pyrimidines Lead to the Identification of a Highly Selective Casein Kinase 2 Inhibitor. *Eur. J. Med. Chem.* 2020, 208, 112770.

(32) Tigno-Aranjuez, J. T.; Benderitter, P.; Rombouts, F.; Deroose, F.; Bai, X.; Mattioli, B.; Cominelli, F.; Pizarro, T. T.; Hoflack, J.; Abbott, D. W. In Vivo Inhibition of RIPK2 Kinase Alleviates Inflammatory Disease. *J. Biol. Chem.* 2014, 289, 29651−29664.

(33) Sánchez-Duffhues, G.; Williams, E.; Benderitter, P.; Orlova, V.; van Wijhe, M.; Garcia de Vinuesa, A.; Kerr, G.; Caradec, J.; Lodder, K.; de Boer, H. C.; Goumans, M. J.; Eekhoff, E. M. W.; Morales-Piga, A.; Bachiller-Corral, J.; Koolwijk, P.; Bullock, A. N.; Hoflack, J.; ten Dijke, P. Development of Macrocycle Kinase Inhibitors for ALK2 Using Fibrodysplasia Ossificans Progressiva-Derived Endothelial Cells. *JBMR Plus* 2019, 3, No. e10230.

(34) Gregg, B. T.; Tymoshenko, D. O.; Razzano, D. A.; Johnson, M. R. Pyrazolo[1,5-a]Pyrimidines. Identification of the Privileged Structure and Combinatorial Synthesis of 3-(Hetero)Arylpyrazolo-[1,5-a]Pyrimidine-6- Carboxamides. *J. Comb. Chem.* 2007, 9, 507−512.

(35) Oliver, A. W.; Knapp, S.; Pearl, L. H. Activation Segment Exchange: A Common Mechanism of Kinase Autophosphorylation? *Trends Biochem. Sci.* 2007, 32, 351−356.

(36) Pike, A. C. W.; Rellos, P.; Niesen, F. H.; Turnbull, A.; Oliver, A. W.; Parker, S. A.; Turk, B. E.; Pearl, L. H.; Knapp, S. Activation Segment Dimerization: A Mechanism for Kinase Autophosphorylation of Non-Consensus Sites. *EMBO J.* 2008, 27, 704−714.

(37) Fedorov, O.; Niesen, F. H.; Knapp, S. Kinase Inhibitor Selectivity Profiling Using Differential Scanning Fluorimetry. *Methods Mol. Biol.* 2012, 795, 109−118.

(38) Wang, Y.; Ma, H. Protein Kinase Profiling Assays: A Technology Review. *Drug Discovery Today: Technol.* 2015, 18, 1−8.

(39) Assay Formats. https://www.reactionbiology.com/services/target-specific-assays/kinase-assays/kinase-assay-services (accessed Nov 16, 2020).

(40) scanMAX Kinase Assay Panel. https://www.discoverx.com/services/drug-discovery-development-services/kinase-profiling/kinomescan/scanmax (accessed Nov 12, 2020).

191

(41) Koeberle, S. C.; Romir, J.; Fischer, S.; Koeberle, A.; Schattel, V.; Albrecht, W.; Grütter, C.; Werz, O.; Rauh, D.; Stehle, T.; Laufer, S. A. Skepinone-L Is a Selective P38 Mitogen-Activated Protein Kinase Inhibitor. *Nat. Chem. Biol.* 2012, 8, 141−143.

(42) Lee, M. R.; Dominguez, C. MAP Kinase P38 Inhibitors: Clinical Results and an Intimate Look at Their Interactions with P38α Protein. *Curr. Med. Chem.* 2005, 12, 2979−2994.

(43) DRAK1/STK17A HotSpot Data sheet. https://reactionbiology.entwicklungsserver.de/datasheet/drak1_stk17a_kin_malvern (accessed Nov 12, 2020).

(44) Müller, S.; Ackloo, S.; Arrowsmith, C. H.; Bauser, M.; Baryza, J. L.; Blagg, J.; Böttcher, J.; Bountra, C.; Brown, P. J.; Bunnage, M. E.; Carter, A. J.; Damerell, D.; Dötsch, V.; Drewry, D. H.; Edwards, A. M.; Edwards, J.; Elkins, J. M.; Fischer, C.; Frye, S. V.; Gollner, A.; Grimshaw, C. E.; Ijzerman, A.; Hanke, T.; Hartung, I. V.; Hitchcock, S.; Howe, T.; Hughes, T. V.; Laufer, S.; Li, V. M.; Liras, S.; Marsden, B. D.; Matsui, H.; Mathias, J.; O'hagan, R. C.; Owen, D. R.; Pande, V.; Rauh, D.; Rosenberg, S. H.; Roth, B. L.; Schneider, N. S.; Scholten, C.; Singh Saikatendu, K.; Simeonov, A.; Takizawa, M.; Tse, C.; Thompson, P. R.; Treiber, D. K.; Viana, A. Y.; Wells, C. I.; Willson, T. M.; Zuercher, W. J.; Knapp, S.; Mueller-Fahrnow, A. Donated Chemical Probes for Open Science. *Elife* 2018, 7, No. e34311.

(45) Wojdyla, J. A.; Kaminski, J. W.; Panepucci, E.; Ebner, S.; Wang, X.; Gabadinho, J.; Wang, M. DA+ Data Acquisition and Analysis Software at the Swiss Light Source Macromolecular Crystallography Beamlines. *J. Synchrotron Radiat.* 2018, 25, 293−303.

(46) McCoy, A. J.; Grosse-Kunstleve, R. W.; Storoni, L. C.; Read, R. J. Likelihood-Enhanced Fast Translation Functions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2005, 61, 458−464.

(47) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2004, 60, 2126−2132.

(48) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 1997, 53, 240−255.

(49) Matulis, D.; Kranz, J. K.; Raymond Salemme, F.; Todd, M. J. Thermodynamic Stability of Carbonic Anhydrase: Measurements of Binding Affinity and Stoichiometry Using ThermoFluor. *Biochemistry* 2005, 44, 5258−5266.

(50) Vasta, J. D.; Corona, C. R.; Wilkinson, J.; Zimprich, C. A.; Hartnett, J. R.; Ingold, M. R.; Zimmerman, K.; Machleidt, T.; Kirkland, T. A.; Huwiler, K. G.; Ohana, R. F.; Slater, M.; Otto, P.; Cong, M.; Wells, C. I.; Berger, B.-T.; Hanke, T.; Glas, C.; Ding, K.; Drewry, D. H.; Huber, K. V. M.; Willson, T. M.; Knapp, S.; Müller, S.; Meisenheimer, P. L.; Fan, F.; Wood, K. V.; Robers, M. B. Quantitative, Wide-Spectrum Kinase Profiling in Live Cells for Assessing the Effect of Cellular ATP on Target Engagement. *Cell Chem. Biol.* 2018, 25, 206−214.

(51) Howarth, A.; Schröder, M.; Montenegro, R. C.; Drewry, D. H.; Sailem, H.; Millar, V.; Müller, S.; Ebner, D. V. HighVia-A Flexible Live-Cell High-Content Screening Pipeline to Assess Cellular Toxicity. *SLAS Discov.: Adv. Sci. Drug Discovery* 2020, 25, 801−811.

(52) Weller, M.; Rieger, J.; Grimmel, C.; Van Meir, E. G.; De Tribolet, N.; Krajewski, S.; Reed, J. C.; von Deimling, A.; Dichgans, J. Predicting chemoresistance in uman malignant glioma cells: the role of molecular genetic analyses. *Int. J. Cancer* 1998, 79, 640−644.

(53) Riss, T. L.; Moravec, R. A.; Niles, A. L.; Duellman, S.; Benink, H. A.; Worzella, T. J.; Minor, L. Assay Guidance Manual [Internet]. *Cell Viability Assays [Updated 2016 Jul 1]*; Markossian, S., Grossman, A., Brimacombe, K., Arkin, M., Auld, D., Austin, C. P., Baell, J., Chung, T. D. Y., Coussens, N. P., Dahlin, J. L., Devanarayan, V., Foley, T. L., Glicksman, M., Hall, M. D., Haas, J. V., Hoare, S. R. J., Inglese, S., Iversen, P. W., Kales, S. C., Lal-Nag, M., Li, Z., McGee, J., McManus, O., Riss, T., Saradjian, P., Sittampalam, G. S., Tarselli, M., Trask, O. J., Jr., Wang, Y., Weidner, J. R., Wildey, M. J., Wilson, K., Xia, M., Xu, X., Eds.; Eli Lilly & Company and the National Center for Advancing Translational Sciences: Bethesda (MD), 2013. https://www.ncbi.nlm.nih.gov/books/NBK144065/.

(54) Ritz, C.; Baty, F.; Streibig, J. C.; Gerhard, D. Dose-Response Analysis Using R. *PLoS One* 2015, 10, No. e0146021.

(55) Van Der Meer, D.; Barthorpe, S.; Yang, W.; Lightfoot, H.; Hall, C.; Gilbert, J.; Francies, H. E.; Garnett, M. J. Cell Model Passports - a Hub for Clinical, Genetic and Functional Datasets of Preclinical Cancer Models. *Nucleic Acids Res.* 2019, 47, D923−D929.

192

Appendix G: Development of a Selective Dual Discoidin Domain Receptor (Ddr)/P38 Kinase Chemical Probe

Reprinted with permission from:

Contribution Menge, A.: Viability assessment using Multiplex high-content live-cell assay, data evaluation, figure preparation, manuscript editing.

Development of a Selective Dual Discoidin Domain Receptor (DDR)/p38 Kinase Chemical Probe

Sandra Röhm, Benedict-Tilman Berger, Martin Schröder, Deep Chatterjee, Sebastian Mathea, Andreas C. Joerger, Daniel M. Pinkas, Joshua C. Bufton, Amelie Tjaden, Lohitesh Kovooru, Mark Kudolo, Christian Pohl, Alex N. Bullock, Susanne Müller, Stefan Laufer, and Stefan Knapp*

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Discoidin domain receptors 1 and 2 (DDR1/2) play a central role in fibrotic disorders, such as renal and pulmonary fibrosis, atherosclerosis, and various forms of cancer. Potent and selective inhibitors, so-called chemical probe compounds, have been developed to study DDR1/2 kinase signaling. However, these inhibitors showed undesired activity on other kinases such as the tyrosine protein kinase receptor TIE or tropomyosin receptor kinases, which are related to angiogenesis and neuronal toxicity. In this study, we optimized our recently published p38 mitogen-activated protein kinase inhibitor 7 toward a potent and cell-active dual DDR/p38 chemical probe and

Dual DDR/p38 inhibitor probe SR-302:
DDR1/2: IC₅₀ = 23/18 nM, p38α/β: IC₅₀ = 125/196 nM

developed a structurally related negative control. The structure-guided design approach used provided insights into the P-loop folding process of p38 and how targeting of non-conserved amino acids modulates inhibitor selectivity. The developed and comprehensively characterized DDR/p38 probe, 30 (SR-302), is a valuable tool for studying the role of DDR kinase in normal physiology and in disease development.
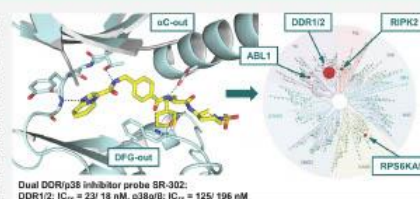
■ INTRODUCTION

The discoidin domain receptor (DDR) kinases DDR1 and DDR2 were first identified in the early 1990s as members of the receptor tyrosine kinase (RTK) family. DDR2 is expressed as a single type-I transmembrane protein, whereas five isoforms of DDR1 produced by alternative splicing have been discovered, namely, DDR1a, DDR1b, DDR1c, DDR1d, and DDR1e.[1,2] DDR kinases share a high similarity in their extracellular discoidin domain structure but differ in their intracellular region. While DDR1a-c and DDR2 share an intracellular kinase domain, the isoforms DDR1d and DDR1e are truncated, kinase-deficient receptors.[3] Typically, RTKs are activated by peptide-like growth factors extracellularly, followed by receptor dimerization, cross-phosphorylation, and fast activation of the signaling cascade. Interestingly, DDR kinases are characterized by a unique mechanism of activation by various forms of triple-helical collagens present in the extracellular matrix.[1,2,4] Furthermore, in contrast to other RTKs, DDRs exist in preformed autoinhibited receptor homo- or heterodimers in the transmembrane, an assembly which enhances collagen binding.[5] Linked to the extracellular matrix, DDR kinases regulate fundamental physiological processes such as cellular adhesion, migration, proliferation, differentiation, matrix remodeling, and survival.[6] Aberrant kinase activation of DDR kinases plays an important role in fibrotic disorders of the lung, kidney, and liver.

DDR1/2 kinases exist mostly in their inactive states if not stimulated by extracellular collagen ligands. The low cellular activity of DDR1 kinase is also caused by an intrinsic autoinhibitory YxxxYY motif, in which the N-terminal tyrosine (Tyr755) controls the autoinhibitory state by phosphorylation.[6] A similar structural motif has previously been described for the insulin receptor tyrosine kinase (IRK), where phosphorylation of a tyrosine reverses autoinhibition and thus enables catalytic activity.[7,8] Another molecular mechanism discussed mediating the unexpected low DDR1 kinase activity inside the cell is a rarely seen salt bridge between Asp671 (αD helix) and Arg752 (A-loop), which stabilizes the inactivating DFG-out conformation.[9] This polar interaction strongly impairs the formation of a catalytically active DFG-in conformation and may explain why crystal structures of DDR1 kinase in complex with type-I inhibitors always show the inactive DFG-out conformation. Interestingly, this stabilizing salt bridge has also been observed in six other kinases, KIT, CSF1R, MEK5, YSK4, and PDGFRα/β, and it has been suggested to play a role in inhibitor
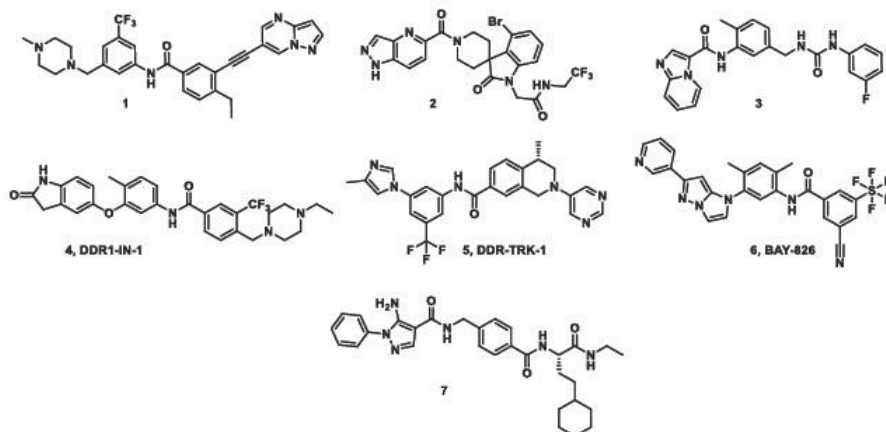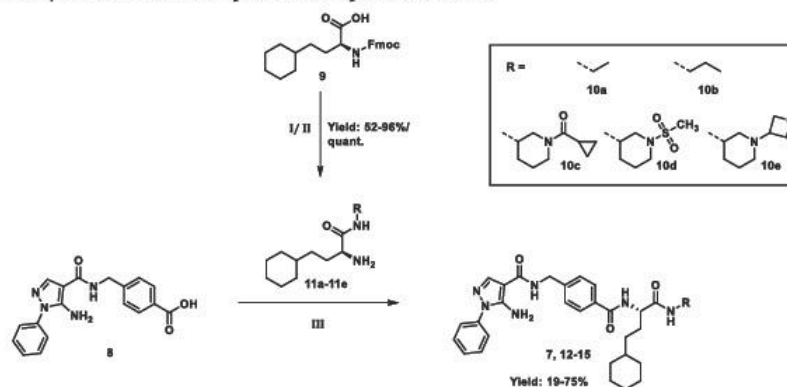
13451

Figure 1. Inhibitor landscape of published DDR1/2 kinase inhibitors (1−6) and the p38 lead compound used for the development of a DDR/p38 chemical probe (7).

Scheme 1. Synthetic Route for the Preparation of Compounds 7 and 12−15[a]

[a]Reagents and conditions: (I) 9, RNH$_2$, HATU, DIPEA, DMF, and RT; (II) piperidine, DMF, and RT; and (III) 11a−11e, HATU, DIPEA, DMF, and RT.

promiscuity.[10] Although DDR kinases have been described as enzymes that bind many diverse inhibitors and selectivity profiles of kinase inhibitors, such as imatinib, frequently show DDR kinases as off-targets, few publications have focused on the design of selective DDR inhibitors (Figure 1).[10] So far, the published inhibitor landscape is of limited scaffold diversity, covering urea compounds designed by derivatization of inhibitors with known DDR1/2 kinase activity, such as imatinib, dasatinib, and BIRB-796.[11−13] For instance, a recent study has focused on ponatinib derivatives, and a diverse set of pyrazolopyrimidine (e.g., 1) and imidazopyrazine derivatives has been evaluated as DDR inhibitors.[14,15] In 2016, Wang et al.

reported tetrahydroisoquinoline-7-carboxamides as very potent and selective DDR1 inhibitors.[16,17] However, these inhibitors, including also the DDR-TRK-1 chemical probe (5),[18] show undesired off-target activities on tropomyosin receptor kinases (TRKA-C), which are important for the normal function of the central and peripheral nervous systems.[16,17,19,20] BAY-826 (6) has been used to probe DDR1/2 kinases, but originally, this inhibitor was developed to study the role of two related RTKs, TIE1/2, in angiogenesis.[21] Roche Pharma discovered a set of very promising potent and selective spiroindolinone derivatives using a DNA-encoded library-screening approach. In this study,

Scheme 2. Synthetic Route for the Preparation of Compounds 20–24[a]
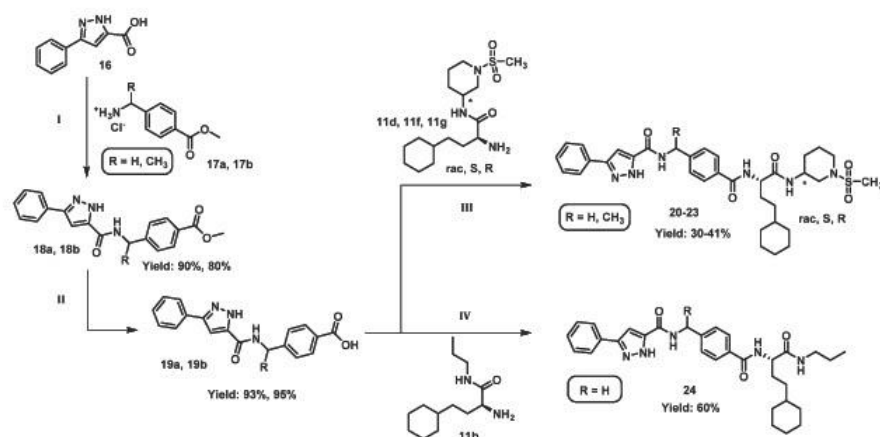
[a]Reagents and conditions: (I) EDC-HCl, HOBt, DIPEA, DMF, and RT; (II) LiOH, THF/H$_2$O, and RT; and (III) 11d, 11f, 11g, HATU, DIPEA, DMF, and RT; and (IV) 11b, HATU, DIPEA, DMF, and RT.

2 prohibited collagen-induced activation of renal epithelial cells expressing DDR1.[22]

To enrich the scaffold diversity of DDR1/2 inhibitors and to enable mechanistic studies on DDR kinase signaling, we optimized compound 7, derived from the p38 inhibitor VPC-00628[23], for targeting DDR1/2 kinases.[24] In our comprehensive characterization of 7, potent activity on DDR1/2 kinases (K_d = 31 nM/40 nM for DDR1/2) was disclosed.[24] Therefore, we chose 7 as a starting point for the synthesis of a comprehensive set of DDR1/2 kinase inhibitors using this chemical scaffold. Structure variations targeting the back pocket and the hinge-binding motif in DDR1/2 kinases provided important structural and mechanistic insights into inhibitor binding. Furthermore, we optimized this series to yield a potent, cell-active dual DDR/p38 chemical probe 30 with excellent selectivity against the kinome and a structurally related negative control.

■ CHEMISTRY

For the synthesis of 7 and its derivatives 12–15, Fmoc-homocyclohexyl-L-alanine (9) was activated with HATU and DIPEA in DMF and reacted with the corresponding amine 10a–10e in an amide coupling reaction. The Fmoc protecting group was then cleaved under basic conditions, and the unnatural amino acids 11a–11e were each reacted with 4-((5-amino-1-phenyl-1H-pyrazole-4-carboxamido)methyl)benzoic acid (8), HATU, DIPEA, and DMF to obtain the final compounds 7 and 12–15 in yields between 19 and 75% (Scheme 1).

For optimizing the hinge-binding head decoration, various heteroaromatic moieties were selected for the synthesis (Schemes 2 and 3). For compounds 20–24, 3-phenyl-1H-pyrazole-5-carboxylic acid (16) was used as a starting material, which was activated with EDC–HCl, HOBt, and DIPEA and then reacted with either benzylamine derivative 17a or 17b

(Scheme 2). The ester functionality was cleaved with lithium hydroxide in a tetrahydrofuran–water mixture to obtain 19a and 19b by precipitation with hydrochloric acid. In the final step, the carboxylic acid was activated by HATU and DIPEA in DMF, and the corresponding amine (11b, 11d, 11f, and 11g) was added, leading to the compounds of interest at yields between 30 and 60%.

A nucleophilic aromatic substitution reaction was applied for the synthesis of compounds 26a and 26b, using the benzylamine derivative 17a or 17b and either triazolopyrimidine 25a or pyrazolopyrimidine 25b as a starting material (Scheme 3). For compounds 26c–26f, an amide coupling reaction was performed with imidazo[1,2-a]pyridine-3-carboxylic acid (25c) or 5-amino-1-methyl-1H-pyrazole-4-carboxylic acid (25d), which were activated with HATU and DIPEA in DMF before the corresponding benzylamine derivative 17a or 17b was added. In the next step, the ester functionality of all compounds synthesized was hydrolyzed with lithium hydroxide to yield the desired acid. Compounds 27a–27f, which were prepared according to this protocol, were then reacted with (S)-2-amino-4-cyclohexyl-N-((SR)-1-(methylsulfonyl)piperidin-3-yl)butanamide (11d) in a second amine coupling reaction to give the final compounds 28–33 at yields between 11 and 65%.

The substituted 3-amino piperidine derivatives used in these amide coupling reactions were synthesized from racemic and/or enantiomeric pure (R), (S) tert-butyl piperidine-3-ylcarbamate (34a–34c, Scheme 4). For the sulfonamide decoration, each of the different 3-amino piperidines 34a–34c was reacted with methanesulfonyl chloride (35a) in a nucleophilic substitution reaction to obtain compounds 36a–36c, respectively. The cyclopropyl decoration was introduced with cyclopropane carboxylic acid (35b), which was reacted with rac 3-amino piperidine (34a) in an HATU-assisted amide coupling reaction. For the oxetane derivative 36e, a reductive amination reaction

Scheme 3. Synthetic Route for the Preparation of Compounds 28−33[a]
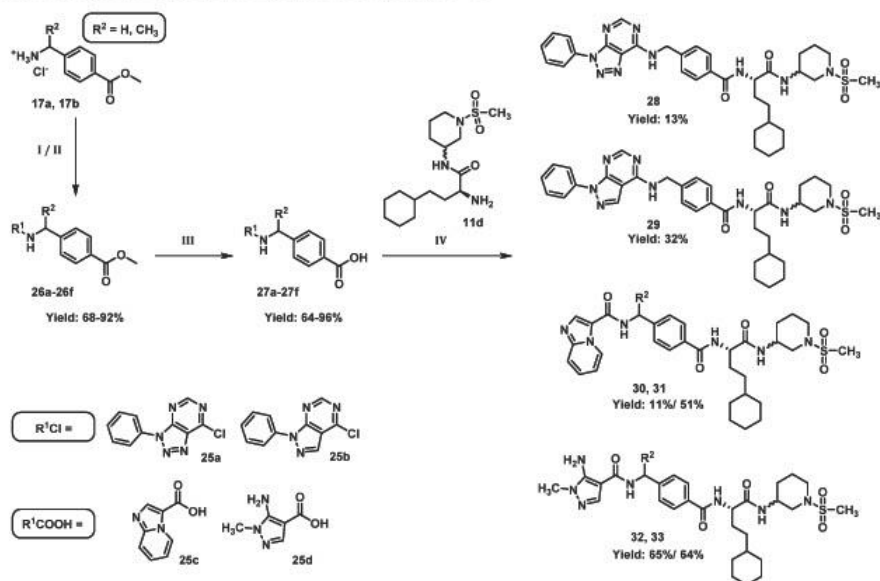
[a]Reagents and conditions: (I) for nucleophilic substitution: DIPEA, THF, and 60 °C; (II) for amide coupling: EDC-HCl, HOBt, DIPEA, DMF, and RT; and (III) LiOH, THF/H$_2$O, and RT; and (IV) HATU, DIPEA, DMF, and RT.

Scheme 4. Synthetic Route for the Preparation of Amine Intermediates (37a−37e)[a]
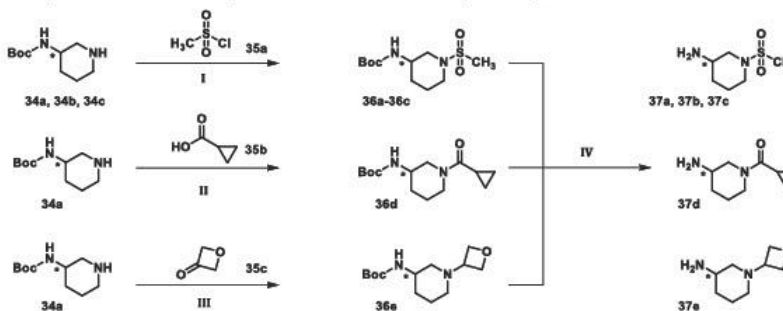
[a]Reagents and conditions: (I) TEA, DCM, and −20 °C → RT; (II) HATU, DIPEA, DMF, and RT; (III) NaBH(OAc)$_3$, HOAc, DCM, and RT; and (IV) TFA, DCM, and 0 °C → RT.

with oxetane-3-one (34c) and sodium triacetoxyborohydride under acidic conditions was performed. The Boc protecting group of all compounds synthesized was removed with trifluoroacetic acid to quantitatively obtain the free amines 37a−37e.

■ **RESULTS AND DISCUSSION**

In order to obtain initial data on the most suitable building blocks for the development of a DDR inhibitor, our previously published p38 compound library covering 127 inhibitors was screened by differential scanning fluorimetry (DSF) against the most potent targets found in the selectivity profiling of 7.[24] The

13454
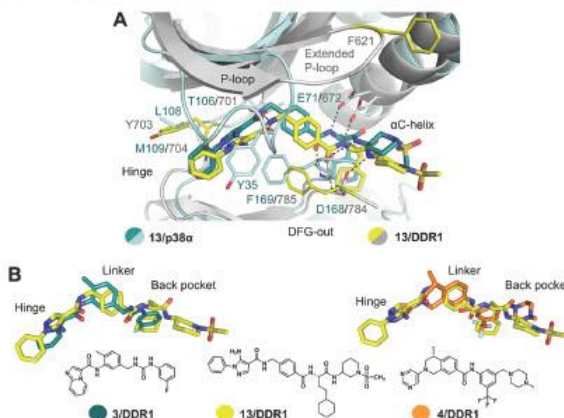
Figure 2. Differences in the binding mode of 13 in p38α and DDR1 and shape similarities with published DDR1 kinase inhibitors. (A) Structure alignment of 13/p38α (light blue, inhibitor in teal, PDB: 7BE4) and 13/DDR1 kinase complex (gray, inhibitor in yellow, PDB: 7BE6). Hydrogen bonds are highlighted as black dashed lines; important structural motifs are labeled (teal for DDR1 and gray for p38α). (B) Superimposed binding modes of 13 (yellow, PDB: 7BE6), compound 3 (teal, PDB: 5BVN) published by Astex and the DDR-TRK-1 precursor 4 (orange, PDB: 5FDX).

p38 compound library contained VPC-00628 derivatives,[25] a set of back-pocket-modified type-II inhibitors and allosteric fragments derived from BIRB-796.[24] Thermal shift values (ΔT_m) were determined for DDR1/2, abelson tyrosine protein kinase 1 (ABL1), and mitogen-activated protein kinase kinase kinase 20 (ZAK) and compared with those of p38α/β (Supporting Information Tables S1 and S2). Because the ΔT_m values typically correlate well with K_d values, potencies of new derivatives were estimated by comparing the corresponding thermal shift values. Compound 7 with known K_d values was used as a control in these experiments (p38α/β [K_d = 6.3/20 nM], DDR1/2 [K_d = 31/40 nM], ZAK [K_d = 120 nM], and ABL1 [K_d = 130 nM]). Comparing the screening data for this compound collection[24] revealed that thermal stabilization of DDR1/2 followed ΔT_m values measured for p38α, despite the low sequence homology of these two kinases. A similar behavior was found for the off-targets ZAK and ABL1. The highest temperature shifts for DDR1 (ΔT_m > 4 °C) were measured for compounds 7 and 13−15. The aminopiperidine back-pocket decorations of compounds 13−15 were therefore considered as the most useful starting point for the synthesis of new DDR1/2 inhibitors. For our VPC-00628[23,25]-based library compounds, the highest thermal shifts for DDR1 (ΔT_m > 4 °C) were found for VPC-00628, MCP-040, MCP-042, MCP-047, MCP-074, and MCP-081 (structures depicted in Supporting Information Tables S1 and S2). In this screen, VPC-00628 and MCP-074 showed off-target activities similar to 7 on all five targets investigated. Besides DDR1, MCP-040 also displayed thermal stabilization of ZAK and ABL1. Compound MCP-042 led to a robust stabilization of DDR1 kinase (ΔT_m = 4.3 °C), but it also showed the highest thermal shift for the off-target ZAK (ΔT_m = 6.5 °C) in the series. The affinity for p38α/β was slightly reduced in comparison with 7, which may be due to the CF₃-decorated pyrazole hinge-binding motif. The latter plays an important role in stabilizing the P-loop in a specific conformation, as confirmed by crystal structure determination
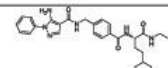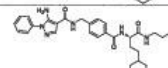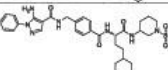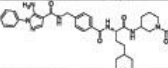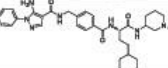
of MCP-042 in complex with p38α.[25] Interestingly, MCP-047 was the only compound that showed preferential stabilization of DDR1/2 (ΔT_m = 4.4 °C), with negligible temperature shifts for all other kinase targets investigated. MCP-047 and VPC-00628 have a similar back-pocket decoration, but they differ in their hinge-binding moiety, which is a 3-phenyl-1H-pyrazole-5-carboxamide in the case of MCP-047. This building block might therefore be useful as a hinge-binding moiety to optimize 7 for the inhibition of DDR1/2 kinase. The chimera of VPC-00628[23] and BIRB-796, MCP-081, previously published and discussed because of its increased α-isoform selectivity,[24,25] showed a slight preference for DDR1 kinase in the thermal shift assays (ΔT_m = 5.3 °C). However, BIRB-like fragments were evaluated in DSF assays (Supporting Information Table S1) and appeared not to be optimal for improving activity toward DDR1/2. We speculated that the sulfonamide-, cyclopropyl amide-, or oxetane-decorated 3-amino piperidines 13−15, respectively, might be better suited for targeting the back pocket of DDR1/2 kinase.

Next, we crystallized 13 in complexes with DDR1 and p38α kinase, respectively, to compare binding modes and to reveal potential sites for DDR1 kinase selectivity (Figure 2A).

Although both kinases belong to different kinase groups (DDR1 is a RTK and p38α is a CMGC family member) and thus share low sequence identity, key interactions with the inhibitor were, however, conserved, including the canonical hinge backbone interactions and polar interactions with the conserved αC glutamate. While 13 engaged the P-loop inside the ATP-binding pocket in p38α via interactions with Tyr35, an extended P-loop conformation was observed in the DDR1 complex. In both structures, inhibitor binding stabilized a (DFG-out) type-II binding mode, with the cyclohexyl moiety protruding into the allosteric DFG-out pocket. Only weak electron density was observed for the sulfonamide moiety, suggesting that it is flexible and does not form specific interactions with the protein or that it adopts alternative binding modes. The hinge region seemed to
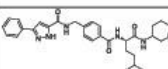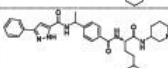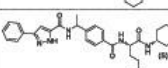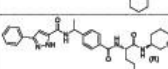
13455
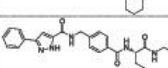
Table 1. Back-Pocket Optimization[a]

No.	Structure	ΔT_m [°C] p38α	ΔT_m [°C] DDR1	ΔT_m [°C] DDR2	IC$_{50}$ [μM] p38α	IC$_{50}$ [μM] DDR1	IC$_{50}$ [μM] DDR2
7		16.4 ± 0.2	4.4 ± 0.1	5.1 ± 0.4	0.014 ± 0.006[g]	4.19 ± 0.84	4.00 ± 0.72
12		13.7 ± 0.5	2.8 ± 0.2	3.8 ± 0.3	0.017 ± 0.003[e]	13.6[d]	5.25 ± 2.84[e]
13		14.1 ± 0.4	4.6 ± 0.2	4.8 ± 0.4	0.033 ± 0.002[f]	2.37 ± 0.28[f]	3.46 ± 0.32[f]
14		14.7 ± 0.5	4.6 ± 0.2	5.4 ± 0.2	0.092 ± 0.004	1.31 ± 0.12[f]	1.75 ± 0.10[f]
15		14.4 ± 0.2	4.9 ± 0.1	5.8 ± 0.0	0.057 ± 0.002	2.43 ± 0.66[e]	3.72 ± 0.30[e]

[a]Potency of compounds 7 and 12−15 as determined by DSF and *in cellulo* NanoBRET assays for p38α and DDR1/2. [b]ΔT_m average ± SD derived from seven replicates for p38α and three replicates for DDR1/DDR2 at a compound concentration of 10 μM. [c]Cellular IC$_{50}$ values determined by NanoBRET assay in HEK293T cells; IC$_{50}$ values ± SEM were derived from technical replicates (n = 3). [d]n = 1. [e]n = 2. [f]n = 4. [g]n = 6.

Table 2. Hinge-Binder Optimization A[a,c]

No.	Structure	ΔT_m [°C] p38α	ΔT_m [°C] DDR1	ΔT_m [°C] DDR2	IC$_{50}$ [μM] p38α	IC$_{50}$ [μM] DDR1	IC$_{50}$ [μM] DDR2
20		1.9 ± 0.4	1.3 ± 0.4	2.0 ± 0.1	12.5 ± 5.8	39.4 ± 19.2	21.5 ± 2.52
21		-0.3 ± 0.2	1.7 ± 0.2	1.7 ± 0.0	18.9 ± 0.9	16.5[d]	27.7 ± 11.3
22		0.4 ± 0.3	1.1 ± 0.2	1.5 ± 0.0	6.0 ± 2.5	45.2 ± 17.9	22.0 ± 7.10
23		0.0 ± 0.3	1.7 ± 0.2	1.9 ± 0.1	6.9 ± 0.5	13.2[d]	35.2 ± 9.4
24		1.0 ± 0.3	0.1 ± 0.3	1.1 ± 0.1	7.2 ± 0.4	10.4[d]	59.2 ± 35.8

[a]Potency of compounds 20−24 determined by ΔT_m and *in cellulo* IC$_{50}$ values for p38α and DDR1/2. [b]ΔT_m average ± SD derived from seven replicates for p38α and three replicates for DDR1/DDR2 at a compound concentration of 10 μM. [c]Cellular IC$_{50}$ values ± SEM determined by NanoBRET assay in HEK293T cells; IC$_{50}$ values were derived from technical duplicates (n = 2). [d]n = 1. [e]n = 3.

be an attractive target for improving inhibitor activity for DDR1/2 because of its high degree of sequence diversity. To identify the most promising hinge-binding moieties, all inhibitors crystallized in complex with DDR1 kinase and published in the KLIFS database (https://klifs.vu-compmedchem.nl/) were compared with the structure in complex with 13 to potentially reveal differences in inhibitor binding modes. The most similar compound to our series was 3

199

Table 3. Hinge-Binder Optimization B[a]

No.	Structure	ΔT_m[b] [°C] p38α	ΔT_m[b] [°C] DDR1	ΔT_m[b] [°C] DDR2	IC$_{50}$[c] [μM] p38α	IC$_{50}$[c] [μM] DDR1	IC$_{50}$[c] [μM] DDR2
13		14.1 ± 0.4	4.6 ± 0.2	4.8 ± 0.4	0.033 ± 0.002[e]	2.37 ± 0.28[e]	3.46 ± 0.32[e]
28		0.5 ± 0.5	0.3 ± 0.1	0.6 ± 0.0	5.39 ± 3.10	21.9 ± 5.44	15.5 ± 5.08
29		6.6 ± 0.3	5.7 ± 0.7	6.1 ± 0.7	0.24 ± 0.02	1.04 ± 0.17[e]	2.06 ± 0.27[e]
30		11.1 ± 0.5	7.1 ± 0.1	9.1 ± 0.6	0.125 ± 0.011[d]	0.023 ± 0.002	0.018 ± 0.002
31		n.d.	7.3 ± 0.3	7.9 ± 0.3	0.284 ± 0.007[d]	0.044 ± 0.001[d]	0.032 ± 0.005[d]
32		9.8 ± 0.3	6.9 ± 0.2	8.7 ± 0.1	1.93 ± 0.05	1.33 ± 0.27	1.16 ± 0.07
33		8.7 ± 0.2	6.1 ± 0.5	6.7 ± 0.5	2.97 ± 0.06	2.27 ± 0.16	7.46 ± 2.38

[a]Potency of compounds 13 and 27−32 for p38α and DDR1/2 determined by DSF and NanoBRET assays. [b]ΔT_m average ± SD derived from seven replicates for p38α and three replicates for DDR1/DDR2 at a compound concentration of 10 μM. [c]Cellular IC$_{50}$ values ± SEM determined by NanoBRET assays in HEK293T cells; IC$_{50}$ values were derived from triplicates ($n = 3$). [d]$n = 2$. [e]$n = 4$.

from Astex Pharmaceuticals[12] and the DDR-TRK-1 precursor 4 (Figure 2B). Although derived from a completely different scaffold, 3 showed conformational similarity with 13 and its interactions with the hinge and DFG-out pocket regions. Furthermore, 3 had a lower molecular weight and did not contain an allosteric, solvent-exposed sulfonamide decoration. An imidazo[1,2-a]pyridine hinge-binding motif was used in 3, forming polar contacts with Met704. The flat geometry of the heterocycle enabled weak face-to-face π-interactions with Tyr703. For 13, an orthogonal geometry between the pyrazole-phenyl decorations was observed, which partially extended into the front pocket. Thus, changing the hinge-binding head group of 13 to an imidazo[1,2-a]pyridine moiety might improve potency toward the DDR1 kinase. The alignment of DDR1-TRK-1 precursors 4 and 13 highlighted shape similarities in the allosteric DFG-out pocket region and the solvent-exposed region. Although the 1,2,3,4-tetrahydroisoquinoline moiety in 4 is more rigid, both chemical scaffolds adopted a similar geometry in the linker and hinge regions. The DDR1-TRK-1 precursor 4 formed hydrogen bonds with Met704 via its pyrimidine heterocycle nitrogen atom. Furthermore, both database-derived compounds harbored a methyl decoration in the linker part that targeted the hydrophobic pocket behind the gatekeeper residue Thr701. We, therefore, speculated that the

introduction of a methyl group to the benzyl linker of 13 may also improve DDR1/2 kinase activity.

To validate the ΔT_m values determined for the inhibitors targeting the allosteric back pocket, in cellulo IC$_{50}$ values of compounds 7 and 12−15 for p38α and DDR1 and DDR2 kinase were determined (Table 1). Although nanomolar potencies were measured for p38α, IC$_{50}$ values for DDR1 and DDR2 were collectively higher in the cellular environment. For 7, an approximately >100 times reduced cellular potency on DDR1/2 was determined compared with the K_d values on the isolated enzyme. Differences in potency observed intracellularly compared with K_d values determined on isolated enzymes were not surprising, given that NanoBRET assays usually measure kinase binding in the full-length proteins at cellular ATP concentrations.

Although the potency for p38α kinase was largely retained for compound 12, potencies for DDR1/2 kinases increased 3.2-fold and 1.3-fold. As expected from the very similar ΔT_m values obtained for the sulfonamide-, cyclopropanamide-, and oxetane-decorated amino piperidines, low-nanomolar affinity binding of p38α was determined. While 13 and 15 inhibited both DDR1/2 kinases with micromolar potencies, 14 showed the best potency in this series, with IC$_{50}$ values of 1.31 and 1.75 μM for DDR1/2 kinase, respectively.

Because the previously used hinge-binding motif seemed to be less optimal for DDR1/2 inhibition, further heterocyclic compounds were evaluated (Table 2). The sulfonamide decoration was left unchanged, and the hinge-binding moiety of 13 was replaced with a 3-phenyl-1H-pyrazole isomer (compounds 20–24). This structural element has already been used for the synthesis of VPC-00628 derivatives,[25] and DSF assays with 7 (Supporting Information Table S2) showed that this heterocycle in compound MCP-047 resulted in better thermal stabilization of DDR1. In order to find out whether the exact stereochemistry of the racemic aminopiperidine derivative in 13 played a role in the binding to DDR kinases, enantiomerically pure (S)- and (R)-configured derivatives were synthesized and investigated. Known DDR kinase inhibitors, such as compounds 3, DDR-1-IN-1, DDR-1-IN-2, and the DDR-TRK-1 probe, all contain a methyl group targeting the hydrophobic pocket located adjacent to the gatekeeper. To investigate the role of the methyl group, a methyl group was incorporated into the benzyl linker of compounds 21–23. In compound 24, the sulfonamide decoration was finally replaced with a propyl residue to investigate the contribution of 3-phenyl-1H-pyrazole as a hinge binder and to exclude combinatorial mismatch effects.

Consistent with the data presented in Supporting Information Table S2 and in our previous publication,[25] the introduction of a 3-phenyl-1H-pyrazole decoration prevented binding to p38α. However, micromolar potencies and negligible thermal shift values were determined also for DDR1/2 kinases. Regarding the role of the configuration of the amino piperidines, very similar binding potencies were obtained for both stereoisomers, highlighting that the stereochemistry at this particular position does not affect inhibitor potency. These observations are in agreement with the crystal structure of 13 in complex with DDR1 kinase, showing that the (S)- and (R)-configurated aminopiperidine moiety pointed toward the solvent. Comparing the ΔT_m and IC_{50} values determined for all kinase targets of 21 (with a methyl group) and 20 (without a methyl-decorated benzyl linker), no significant difference in binding potency was apparent. Due to the unfavorable 3-phenyl-1H-pyrazole hinge-binding building block, only negligible inhibitory effects were determined for compound 24 with a propyl back-pocket decoration.

Next, we focused on the decoration of the aminopyrazole moiety in 13 (Table 3). In 32 and 33, the phenyl moiety thought to be crucial for P-loop stabilization in p38 MAPK and, hence, inhibitor selectivity was truncated. Thermal stabilization of p38α with 32 ($\Delta T_m = 9.8$ °C) and 33 ($\Delta T_m = 8.7$ °C) was significantly lower than that with 13 ($\Delta T_m = 14.1$ °C), supporting our hypothesis. Interestingly, thermal stabilization of DDR1/2 kinases was even higher for the phenyl-truncated compounds, which was confirmed by BRET assays. While 32 and 33 showed about 58-fold and 90-fold reduction in potency for p38α compared with 13, IC_{50} values of 1.33 and 1.16 μM were determined for 32 for DDR1/DDR2 kinases. Compound 33 had slightly higher IC_{50} values for all kinase targets, presumably due to unfavorable contacts of the methyl-stabilized benzyl linker. To study if the reduced affinity for p38 MAPK is caused by altered P-loop folding/interactions, a crystal structure of 32 in complex with p38α kinase was determined (Figure 3).

When comparing the crystal structures of p38α in complex with compounds 13 and 32, remarkable differences in the P-loop conformation became apparent. While 13 engaged the P-loop via its phenyl residue by a face-to-face stacking interaction with
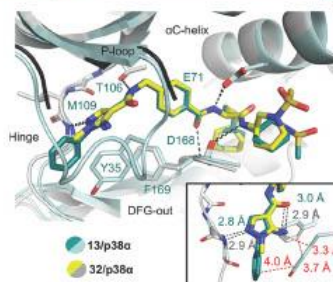
Figure 3. Mechanism of P-loop folding in p38 MAPK. Alignment of 13 (teal, PDB: 7BE4) with 32 (yellow, PDB: 7BE5) in complex with p38α. Compound 13 engages the P-loop in a folded conformation (teal). With the phenyl-deficient compound 32, the P-loop was disordered and could not be modeled. Hydrogen bonds are highlighted with black dashed lines. Inset: stabilization of the P-loop with 13 with shown hydrogen bonds (black dashed lines) and distance measurements (red dashed lines). Important structural motifs are labeled.

Tyr35 and side-on stacking interaction with Phe169 from the DFG motif, the P-loop was disordered in the complex with 32. This observation confirmed the hypothesis that the p38α hinge-targeting phenyl moiety of 32 was crucial for inhibitor potency and for trapping the P-loop inside the ATP-binding pocket. During the development of the p38α/β type-II probe SR-318 and its negative control compound SR-321, the free amine at the pyrazole ring system was also found to be important for the P-loop interaction,[25] but the structural analysis mentioned above suggests that the phenyl-mediated interactions are the driving force for stabilizing the P-loop conformation.

Analysis of the crystal structures revealed that an intra-molecular hydrogen bond was formed between the pyrazole-free amine group and the neighboring carbonyl oxygen atom. This hydrogen bond stabilized the heterocycle and the amide group in a planar conformation, facilitating hinge-binding interaction with Met109 and the gatekeeper Thr106. Bridging of both groups was therefore considered. Compounds 29 and 28 were successfully prepared by the use of a 1-phenyl-1H-pyrazolo[3,4-d]pyrimidine or 3-phenyl-3H-[1,2,3]triazolo[4,5-d]pyrimidine building block. ΔT_m and IC_{50} values were determined for both compounds, giving somewhat unexpected results. For compound 29 [IC_{50} (DDR1/2) = 1.04/2.06 μM], the potency toward DDR1/2 was improved compared with 13. However, affinity for p38α remained unaffected, which was probably caused by the phenyl decoration in the hinge region. Interestingly, for 28 binding to p38α, DDR1, and DDR2, it was impaired, and only negligible thermal shift values and potencies in NanoBRET assays were determined. The latter finding was unexpected, as only a single carbon atom of the pyrazole was replaced with a nitrogen in the triazolo-decoration. These data suggested that the triazolo[4,5-d]pyrimidine in 28 impaired binding due to altered electrostatic complementarity. Compound 28 might therefore be used as a potential negative control compound.

In compounds 30 and 31, the hinge-binding heterocycle was replaced with a bicyclic imidazo[1,2-a]pyridine decoration, reminiscent of compound 3 for targeting DDR1/2 kinase. Gratifyingly, the measured IC_{50} values of 30 and 31 for DDR1/2

kinases were significantly lower than IC_{50} values of earlier compounds. For 30, IC_{50} values of 23 and 18 nM were determined for DDR1 and DDR2, respectively. For compound 31, harboring a methyl-stabilized benzyl linker, potencies were only slightly reduced. However, activity for p38α kinase remained in both inhibitors, 30 and 31, with potencies 5−6 times higher for p38α than for DDR1.

To understand the differences in inhibitor binding in p38α and DDR1 kinase, a crystal structure in complex with 30 was determined for both targets (Figure 4).
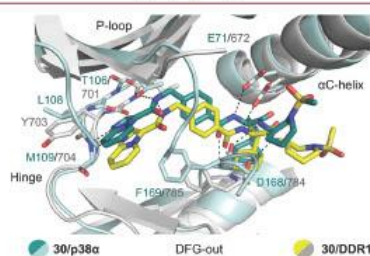
Figure 4. Binding mode of dual DDR/p38 inhibitor 30. The crystal structure of DDR1 kinase in complex with 30 (yellow inhibitor, PDB: 7BCM) showing an extended P-loop. In contrast, in the p38α kinase complex with 30 (teal inhibitor, PDB: 7BDO), the P-loop adopts a folded conformation where the ligand interacts with the P-loop (light blue). In both cases, the tip of the P-loop was partly disordered, however. Hydrogen bonds are highlighted as black dashed lines; important structural motifs are labeled.

Overall, 30 adopted a similar binding mode in both kinases, targeting an inactive DFG-out state, with the cyclohexyl decoration interacting with the allosteric deep pocket, which is slightly smaller in DDR1 kinase. The sulfonamide moiety pointed toward the solvent. In both structures, a similar hydrogen network was seen engaging the Glu−Asp salt bridge of the DFG motif and the conserved glutamate in αC, and an additional hydrogen bond formed with the DFG aspartate and the terminal amide. However, differences were observed in interactions with the hinge region and the P-loop. While in p38α, Met109 is preceded by a leucine (Leu108), an aromatic tyrosine residue (Tyr703) is found in DDR1 at this position. The side chain of Tyr703 pointed toward the ATP-binding pocket and interacted with the imidazo[1,2-a]pyridine hinge-binding decoration. The flat geometry of this hinge binder was probably the main driver for the improved potency of 30 for DDR1/2 kinase. In the DDR1−13 complex (Figure 2A), an orthogonal orientation of aminopyrazole and its phenyl decoration were found, negatively influencing aromatic interactions with Tyr703. In addition to the potency toward DDR1/2 kinase, 30 also showed potent binding of p38α/β. The reason for this can also be explained based on the crystal structure, which showed that 30 traps the P-loop of p38α in a conformation that forms a lid over the inhibitor binding site, reminiscent of the folded P-loop seen with 13. However, not all residues in the P-loop region were fully resolved in the crystal structure, and π-stacking interactions with Tyr35 and Phe169 from the DFG motif were not visible in this structure, indicating that the P-loop retained a certain degree of flexibility. As discussed above, 13 showed a very low IC_{50} value of 22.5 nM for p38α, probably mainly due to favorable interactions with the P-loop, which was fully resolved in the crystal structure (Figure 2A). As the hinge decoration of 30 lacked the phenyl-extended hinge-binding head group, interactions with the P-loop Tyr35 were less pronounced, which could explain the 4.8-fold reduced potency for p38α.

Selectivity Profiling. The most interesting and potent compounds were selected to study their kinome-wide selectivity. Therefore, 28, 29, and 30 were each profiled against 468 kinases
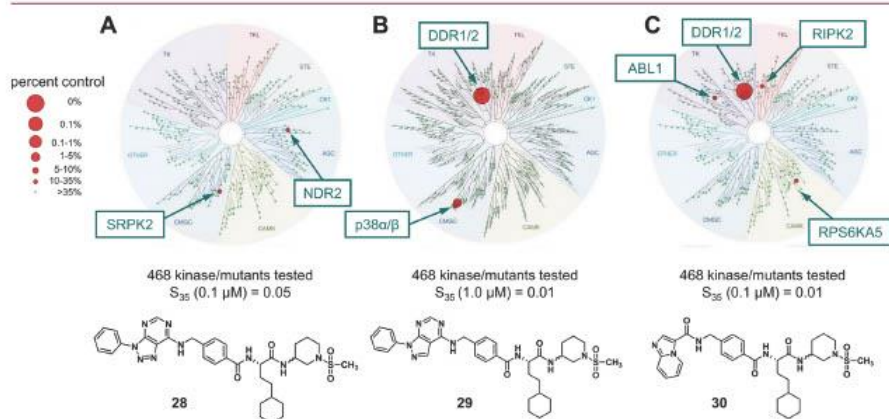
Figure 5. Selectivity profiling against 468 kinases and kinase mutants. (A) 28 screened at 0.1 μM; (B) 29 screened at 1.0 μM; and (C) 30 screened at 0.1 μM; the selectivity score S_{35} is given for compound comparison; mutants not shown; the selectivity screen was performed using Eurofins scanMAX technology.

and clinically relevant kinase mutants in total (Figure 5, Supporting Information Table S6). All compounds showed very clean selectivity profiles, as expected based on earlier studies on structurally related compounds.

Compound 29 which contained a bridged pyridazo[3,4-d]pyrimidine structure was screened at a concentration of 1 μM (Figure 5B), and a remarkably clean selectivity profile was obtained. From the 468 kinases and mutations screened with 29, only DDR1/2 and p38α/β kinases were inhibited by more than 50% [selectivity score of S_{35} (1 μM) = 0.01]. To understand the reason for this kinome-wide selectivity, a crystal structure in complex with p38α was determined (Figure 6).
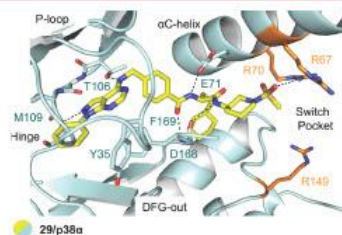
Figure 6. Selectivity by targeting the switch-pocket residues in p38α. Compound 29 (teal, PDB: 7BDQ) in complex with p38α MAPK. Important structural elements are highlighted, including the switch-pocket residues Arg70, Arg67, and Arg149 (shown in orange). The structure reveals a folded P-loop conformation stabilized by inhibitor binding. Hydrogen bonds between the inhibitor and the kinase are highlighted by black dashed lines.

In this crystal structure, 29 bound as expected as a typical type-II inhibitor occupying the DFG-out deep pocket and made contacts with the backbone amide of Met109 in the hinge region via its pyrazole nitrogen atom. Besides this canonical binding mode, 29 engaged the P-loop in a folded conformation that acted as a lid over the hinge-binding moiety, which improved inhibitor selectivity, as seen for other inhibitors crystallized in this study. However, none of these inhibitors matched the excellent kinome-wide selectivity of 29. The reason for this selectivity might be due to an unexpected hydrogen bond formed between Arg70 from the αC helix and one of the sulfonamide O-atoms, which was not seen in the p38 complexes with 13, 32, and 30. Like Arg67 (αC helix) and Arg149 (HRD segment), Arg70 belongs to the so-called switch-pocket residues. Targeting these amino acids was previously studied by Deciphera Pharmaceuticals, who developed highly selective p38α/β inhibitors.[26] Since Arg70 is a non-conserved amino acid found exclusively in p38α/β, the superior kinome-wide selectivity of 29 may be primarily caused by the additional interaction with the basic arginine side chain. The binding mode of 29 was unique, as it represented the first switch-pocket inhibitor simultaneously making contacts with the hinge region. Another unique feature was that it trapped the P-loop in an unusual distorted conformation.

The most potent compound of our structure−activity relationship (SAR) series, 30, was screened against the kinome at a concentration of 0.1 μM to obtain insights into its selectivity (Figure 5C). In addition to DDR1/2, which were bound the strongest, three other kinases were detected with a residual activity of <35% of the uninhibited control at this concentration:

ABL1, RIPK2, and RPS6KA5 (Table 4). The cellular activity of these targets and p38α/β was studied by NanoBRET and

Table 4. Selectivity Profiling and Target Activity of 30

| kinome targets	% activity compared with the uninhibited control at 0.1 μM	NanoBRET IC$_{50}$ [μM][a]
DDR2	0	0.018 ± 0.002
DDR1	0.1	0.023 ± 0.002
ABL1-nonphosphorylated	14	>25 μM
RIPK2	30	>25 μM
RPS6KA5 (Kin.Dom.2-C-terminal)	35	>25 μM
p38β	36	0.196 ± 0.008
p38α	77	0.125 ± 0.011

[a]IC$_{50}$ values ± SEM were derived from triplicates (n = 3).

revealed potent inhibitory potencies of 30 on DDR1/2 (IC$_{50}$ = 23/18 nM) and p38α/β (IC$_{50}$ = 125/196 nM) kinases (Table 4). A significantly lower cellular activity was found for the additional targets, with IC$_{50}$ values >25 μM, making 30 a very valuable tool compound (Supporting Information Figure S1, Tables S3 and S4).

Compound 28, which had no significant inhibitory activity on p38α and DDR1/2 in the DSF screen, was screened at a concentration of 0.1 μM (Figure 5A). Interestingly, 28 also showed a very clean selectivity profile in the kinome scan assay. With a cutoff value of >35% activity of the control in the assay, only two kinases, SRPK2 and NDR2, were detected [S_{35} (0.1 μM) = 0.005]. Compared with the potent inhibitor 30, compound 28 showed more than 850-fold reduced potency for DDR1/2 kinases (IC$_{50}$ = 21.9/15.5 μM), a 43/8-fold decreased IC$_{50}$ value for p38α/β (IC$_{50}$ = 5.4/1.6 μM), and no additional off-targets. It, therefore, represents a valuable negative control of this compound series for cell-based assays.

In compounds 31 and 33, a methyl group was introduced in the benzylic position to strengthen the interaction with the hydrophobic pocket behind the gatekeeper (Thr701) and to stabilize the compounds metabolically. This methyl group is part of many previously published DDR1/2 kinase inhibitors, such as DDR-TRK-1, DDR1-IN-1, and 3, and it is believed to be responsible for the off-target inhibition of TRKs.[27] Since none of the previously synthesized compounds based on the VPC-00628 scaffold showed any binding of TRKs in Eurofins scanMAX, it was of interest whether a methyl group at this position in the molecule also leads to the aforementioned interaction. Compound 31 was therefore investigated for its kinome-wide selectivity at a concentration of 0.1 μM (Supporting Information Table S6). Interestingly, screening of 31 revealed a clean profile, and high-affinity targets such as DDR1/2, p38α, and MELK were detected at a cutoff value of >35% activity of the control. The anticipated inhibition of TRKA, TRKB, and TRKC based on the scaffold 31 was, however, not confirmed, and 100% activity of the control was retained in the Eurofins scanMAX kinase assay panel. For compound 33, which was studied at the same concentration, a similar clean selectivity profile, without inhibition of TRKs, was observed (Supporting Information Table S6). It can therefore be assumed that the methyl group is not the only reason for the inhibition of TRKs, as discussed by Zhu et al.,[27] but that this activity is rather dependent on the hinge-binding scaffold of the inhibitor used.

Cytotoxicity, TNF-α Release, and Drug-like Properties. We investigated the dose- and time-dependent effects of the
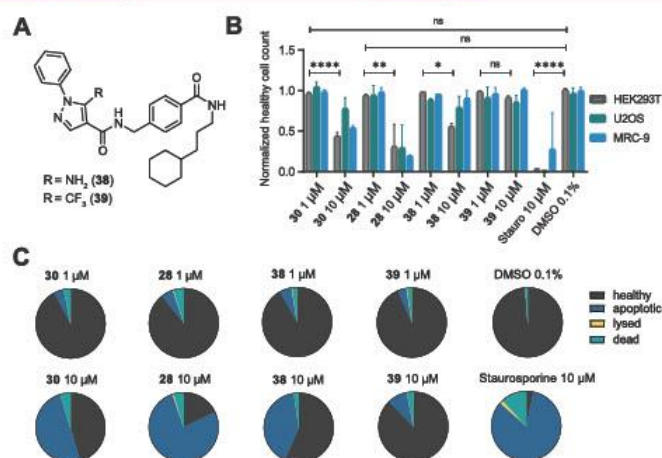
Figure 7. Live cell high-content screen of HEK293T, U2OS, and MRC-9 cells. (A) Structure of recently published p38 probe 38 (SR-318) and control 39 (SR-321). (B) Healthy cell count after 24 h of 1 and 10 μM compound exposure (30, 28, 38, 39, and staurosporine) normalized against healthy cells exposed to DMSO 0.1% in HEK293T, U2OS, and MRC-9 cells. Error bars show SEM of three technical duplicates. The significance was calculated using a two-way ANOVA analysis. (C) Fraction of healthy, apoptotic, lysed, and dead HEK293T cells after 24 h of exposure to 1 and 10 μM compound (30, 28, 38, 39, and staurosporine) compared with a 0.1% DMSO control. Additional cytotoxicity data with all three cell lines are shown in Supporting Information Figures S2–S4.
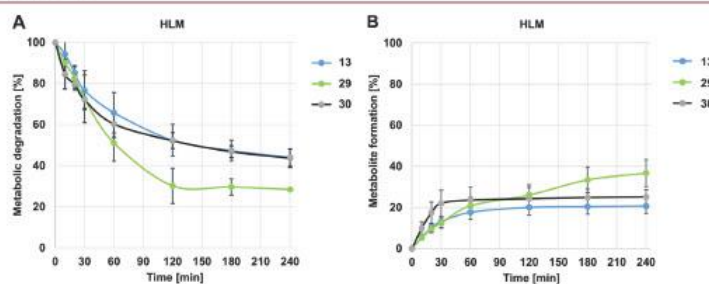
Figure 8. Metabolic stability of compounds 13, 29, and 30 in HLMs. (A) Metabolic degradation over 240 min. (B) Metabolite formation over 240 min. Experiments were performed in triplicate. The values represent the mean with standard deviation.

compounds 28, 30, 38, and 39 on cell viability in three different cell types—human embryonic kidney cells (HEK293T), osteosarcoma cells (U2OS), and non-transformed human lung fibroblasts (MRC-9) (Figure 7 and Supporting Information Figures S2–S4). Collagen-activated DDR1/2 cell signaling is mediated by mitogen-activated protein kinases, such as ERK1/2, JNK, and p38α/β.[28] To exclude p38 off-target cytotoxicity, our previously published p38α/β-selective chemical probe 38 (SR-318) and its control 39 (SR-321) were included in this assay.[25] The cells were treated with fluorescent dyes to detect apoptosis (Alexa Fluor 680 conjugate), nucleus/DNA (Hoechst33342), mitochondrion (MitoTracker red), and microtubule (Bio-Tracker 488 Green Microtubule Cytoskeleton Dye) effects. Fluorescence and cellular shape were measured before compound treatment and 12, 24, and 48 h after compound

exposure using a CQ1 high-content confocal microscope (Yokogawa). Staurosporine at a concentration of 10 μM was used as a reference compound.

After 24 h, the DDR/p38α/β probe compound, 30, and the corresponding negative control (28) showed no significant cytotoxicity at a compound concentration of 1 μM. At 10 μM, however, there was a pronounced cytotoxic effect and a drastic increase in the percentage of apoptotic cells in all three cell lines for these two inhibitors compared with the DMSO-treated controls (Figure 7C and Supporting Information Figure S4). Interestingly, 28 was more toxic than 30 at 10 μM, which was also reflected in a higher proportion of apoptotic cells. For the p38 α/β probe 38 (SR-318) and its control 39 (SR-321), there was also no notable cytotoxicity detected after 24 h in all three cell lines at an inhibitor concentration of 1 μM. No major effect

13461

on the microtubule function was observed after treatment with any of the compounds from the two probe sets, and none of the compounds resulted in increased levels of cell lysis at a concentration of 1 μM (Supporting Information Figures S3 and S4). Thus, the developed probe, 30, and its corresponding negative control, 28, represent a versatile toolset to study the DDR1/2 function in normal or disease physiology.

In our NanoBRET target engagement studies, we showed that 30 also inhibited p38α/β with a five/eightfold lower potency compared with DDR1/2. Recent studies indicated that the collagen-activated DDR receptor may upregulate the expression and release of cytokines in macrophages.[29] This process is dependent on NF-κB signaling and has been suggested to be mediated by the p38 MAPK pathway. To assess the phenotypic effect of our dual DDR/p38 probe, 30, we measured the TNF-α release in whole blood. Besides 30, the negative control 28 and our recently published p38 selective probe set (38 and 39) and the p38 inhibitor SB203580 were studied in this assay. Both p38 inhibitors, 38 and SB203580, potently inhibited the TNF-α release with IC$_{50}$ values of 0.17 and 1.82 μM, respectively, while no inhibition was seen for 30 and the control 39 (Supporting Information Figure S5). Based on these results, we concluded that the phenotypic effect of 30 on the p38α/β MAPK pathway is less pronounced.

We next tested the metabolic stability of compounds 13, 29, and 30 using human liver microsomes (HLMs) over a time interval of 240 min (Figure 8). We observed that 29 was metabolized slightly faster than 13 and 30, and a slightly higher proportion of metabolites was thus detected. After a reaction time of 240 min, only 44.0 ± 4.0% of the compound remained for 13, 43.6 ± 4.5% of the compound remained for 30, and 28.4 ± 0.8% of the compound remained for 29. In summary, all derivatives modified for targeting DDR1/2 showed modest metabolic stability.

To more comprehensively assess drug-like properties of 30, ADME properties were predicted using QikProp software (Schrödinger, 2020, Supporting Information Table S9). Overall, 30 fulfilled all criterion limits for drug-like compounds, such as having a good predicted cell permeability in Caco-2 and MDCK cells, which was supported by our nanoBRET assay data. In addition, compound 30 had a good predicted oral absorption of 60% and may therefore show reasonable bioavailability *in vivo*.

■ CONCLUSIONS

A comprehensive set of 127 previously published p38 inhibitors was evaluated by DSF against p38 and the main off-targets of the VPC-0062825-derived lead compound, 7, which led to the identification of suitable chemical starting points for the development of selective DDR1/2 inhibitors. A series of potent DDR1/2 kinase inhibitors were synthesized by targeting the hinge-binding moiety and the back-pocket decorations present in the lead structure 7. Characterization of these compounds and the determination of crystal structures of selected inhibitor complexes highlighted the importance of the P-loop conformation for potent interactions with p38, including a rarely seen π-stacking interaction with the inhibitor. Furthermore, we demonstrated the possibility of targeting the non-conserved switch-pocket residue Arg70 in p38α with a type-II inhibitor to achieve p38 selectivity. Our SAR optimization resulted in a very potent DDR inhibitor, 30 (SR-302), with IC$_{50}$ values of 0.023 and 0.018 μM for DDR1/2 and more than five times lower cellular IC$_{50}$ values for p38α/β (IC$_{50}$ = 0.125/0.196 μM). Compound 30 displayed an excellent kinome-wide selectivity, as

demonstrated in a screen against 468 kinases and mutants at a concentration of 1 μM [S-score (35) = 0.01]. In addition, we synthesized an appropriate negative control compound, 28 (SR-301), based on the same scaffold, for the use of our compound series in cell-based assays. Taken together, our data suggest that the DDR/p38 probe set is a valuable tool for in-depth mechanistic studies of DDR1/2 signaling under physiological and pathophysiological conditions.

■ EXPERIMENTAL SECTION

Chemistry. All reagents and water-free solvents were purchased from commercial suppliers and were used, if not otherwise stated, without further purification, or solvents were dried using standard procedures. Column chromatography on silica was performed on silica 60, 0.04–0.63 mm from Macherey-Nagel GmbH & Co. KG. Fully characterized compounds were chromatographically homogenous. Analytical thin-layer chromatography was carried out using aluminum-backed plates coated with Merck TLC Silicagel 60 F254. Plates were visualized under UV light (at λ 254 and/or 360 nm) and/or with ninhydrin solution. ^{1}H NMR and ^{13}C NMR spectra were recorded either on a Bruker AVANCE 250, 300, 400, and 500 or a Bruker DRX 600 using TMS as an internal standard. The chemical shifts (δ) are reported in ppm and are calibrated against the residual proton peak of the deuterated solvent. J values were recorded in Hz, and multiplicities were expressed using the usual conventions. Mass spectrometry (ESI) was measured on a VG Plattform II spectrometer from Fisons. High-resolution mass spectrometry (FTMS + p MALDI-HRMS) was performed using a MALDI LTQ XL Orbitrap spectrometer from Thermo Scientific. Determination of the compound purity by HPLC was carried out on an Agilent 1260 Infinity II device with a 1260 DAD HS detector (G7117C; 254 nm, 280 nm) and a LC/MSD device (G6125B, ESI pos. 100–1000). The compounds were analyzed on a Poroshell 120 EC-C18 (Agilent, 3 × 150 mm, 2.7 μm) reversed-phase column using 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (ACN) (B) as a mobile phase. The following gradient was used: 0 min 5% B, 2 min 5% B, and 8 min 98% B (flow rate of 0.5 mL/min). UV detection was performed at 254 and 280 nm, and all compounds used for further biological characterizations showed >95% purity, if not otherwise stated. For compound 28, purity was determined using a LC-20A Prominence HPLC instrument from Shimadzu. Separations were performed on a C18 column [Luna 10 μm C18 (2) 100 Å; 250 × 4.6 mm] from Phenomenex using the following gradient profile: 0–2 min 95% B, 2–14 min 95% B, 14–21 min 10% B, and 21 min 95% B. As solvent (A), ACN, ultra-MS-grade, was used, and as solvent (B), MS-grade H$_2$O with 0.1% formic acid at a flow rate of 1 mL/min was used. The detection was carried out with an LCMS-2020 mass spectrometric detector from Shimadzu at a wavelength of 254 and 280 nm.

General Procedure A for Fmoc Cleavage and the Preparation of Amines 11a–11g. Amines 11a–11g used in the amide coupling reaction for general procedure B were freshly prepared directly before coupling to the acid. The corresponding Fmoc protected amine (10a–10g, 1.2 equiv) was dissolved in DMF (12 mL), and piperidine (20%) was added. The mixture was stirred for 1 h at RT, and CH$_2$Cl$_2$ (30 mL) or EtOAc (30 mL) was added. The organic phase was washed five times with water (30 mL) and dried over MgSO$_4$. The solvent was evaporated under reduced pressure, and the crude product (quantitative) was used without further purification.

General Procedure B for Amide Coupling Reaction Using HATU for the Preparation of Final Compounds 12–15, 20–24, and 30–33. The corresponding carboxylic acid (1 equiv) and HATU (1.2 equiv) were dissolved in DMF (8 mL), and DIPEA (1.5 equiv) was added. The solution was stirred at RT for 1.5 h, and a solution of the corresponding amine (11a–11g, 1.2 equiv) in DMF (8 mL) with DIPEA (1.5 equiv) was added. The mixture was stirred for 20 h at RT. CH$_2$Cl$_2$ or EtOAc (50 mL) was added, and the organic phase was washed four times with water (50 mL) and dried over MgSO$_4$. The solvent was evaporated under reduced pressure, and the crude product

was purified by column chromatography on silica (eluent: cyclohexane/EtOAc, CH₂Cl₂/MeOH, or EtOAc/MeOH) if not stated otherwise.

(S)-5-Amino-N-(4-((4-cyclohexyl-1-(ethylamino)-1-oxobutan-2-yl)carbamoyl)benzyl)-1-phenyl-1H-pyrazole-4-carboxamide (7), 4-((5-amino-1-phenyl-1H-pyrazole-4-carboxamido)methyl)benzoic acid (8), and (9H-fluoren-9-yl)methyl (S)-(4-cyclohexyl-1-(ethylamino)-1-oxobutan-2-yl)carbamate (10a) were prepared as previously published.[24]

(9H-Fluoren-9-yl)methyl (S)-(4-Cyclohexyl-1-oxo-1-(propylamino)butan-2-yl)carbamate (10b). (S)-2-((((9H-Fluoren-9-yl)methoxy)carbonyl)amino)-4-cyclohexylbutanoic acid (9, 400 mg, 0.98 mmol) and HATU (448 mg, 1.18 mmol) were dissolved in DMF (dry, 10 mL), and DIPEA (152 mg, 1.18 mmol) was added. The reaction mixture was stirred for 2 h at RT, and propan-1-amine (97 μL, 1.18 mmol) was added. The mixture was stirred for additional 20 h at RT. DCM (50 mL) was added, and the organic phase was washed five times with H₂O (20 mL) and dried over MgSO₄. The solvent was evaporated under reduced pressure, and the crude product was purified by column chromatography on silica (cyclohexane/EtOAc 1:1). Compound 10b was obtained in a yield of 248 mg (0.55 mmol, 56%) as a colorless solid. TLC: R_f = 0.15 (SiO₂, cyclohexane/EtOAc, 1:1). ¹H NMR (500 MHz, CDCl₃, 300 K): δ 7.76 (d, ³J = 7.5 Hz, 2H), 7.57 (d, ³J = 7.5 Hz, 2H), 7.39 (t, ³J = 7.4 Hz, 2H), 7.30 (t, ³J = 7.4 Hz, 2H), 6.13 (s, 1H), 5.48 (d, ³J = 7.9 Hz, 1H), 4.48–4.28 (m, 2H), 4.20 (t, ³J = 7.0 Hz, 1H), 4.12–4.02 (m, 1H), 3.30–3.11 (m, 2H), 1.91–1.76 (m, 1H), 1.73–1.58 (m, 6H), 1.56–1.43 (m, 2H), 1.36–1.06 (m, 6H), 0.94–0.79 (m, 5H) ppm. ¹³C NMR (126 MHz, CDCl₃, 300 K): δ 171.8, 156.3, 143.9, 141.4, 127.9, 127.2, 125.2, 120.1, 67.1, 55.5, 47.2, 41.3, 37.6, 33.4, 33.3, 33.1, 30.4, 26.7, 26.4, 22.9, 11.4 ppm. MS (ESI pos.) m/z (%): 897.62 (84) ([2M + H]⁺ calcd, 897.54), 449.24 (92) ([M + H]⁺ calcd, 449.27), 227.18 (100) ([M − Fmoc + H]⁺ calcd, 227.20).

(9H-Fluoren-9-yl)methyl ((S)-4-Cyclohexyl-1-(((SR)-1-(cyclopropanecarbonyl)piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamate (10c). (S)-2-((((9H-Fluoren-9-yl)methoxy)carbonyl)amino)-4-cyclohexylbutanoic acid (9, 500 mg, 1.23 mmol) and HATU (560 mg, 1.47 mmol) were dissolved in DMF (dry, 10 mL), and DIPEA (190 mg, 1.47 mmol) was added. The reaction mixture was stirred for 1.5 h at RT, and rac-(3-aminopiperidin-1-yl) (cyclopropyl)methanone (37d, 248 mg, 1.47 mmol) dissolved in DMF (dry, 5 mL) was added. The mixture was stirred for additional 20 h at RT. EtOAc (100 mL) was added, and the organic phase was washed four times with H₂O (50 mL) and dried over MgSO₄. The solvent was evaporated under reduced pressure, and the crude product was purified using column chromatography on silica (cyclohexane/EtOAc 1:3 → 1:9). Compound 10c was obtained in a yield of 553 mg (0.992 mmol, 81%), as a colorless solid. TLC: R_f = 0.37 (SiO₂, cyclohexane/EtOAc, 1:3). ¹H NMR (400 MHz, DMSO-d₆, 300 K): (mixture of diastereomeres): δ 7.96–7.80 (m,3H), 7.77–7.68 (m, 2H), 7.46–7.37 (m, 3H), 7.32 (t, ³J = 7.4 Hz, 2H), 4.33–4.13 (m, 2H), 4.09–3.78 (m, 2H), 3.76–3.47 (m, 2H), 3.39–3.05 (m, 2H), 2.00–1.29 (m, 12H), 1.27–1.03 (m, 6H), 0.91–0.77 (m, 2H), 0.76–0.53 (m, 4H) ppm. ¹³C NMR (101 MHz, DMSO-d₆, 300 K): (mixture of diastereomeres): δ 171.0, 170.9, 152.0 (2×), 143.9, 143.8, 140.7 (2×), 138.6, 127.6, 127.0, 125.3, 120.1 (2×), 65.5, 46.6, 36.8, 33.0, 32.9, 32.7, 32.6, 26.1, 25.8 (2×), 10.4, 6.8 ppm. MS (ESI pos.) m/z (%): 580.43 (100) ([M + Na]⁺ calcd, 580.32).

(9H-Fluoren-9-yl)methyl ((S)-4-Cyclohexyl-1-(((SR)-1-(methylsulfonyl)piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamate (10d). (S)-2-((((9H-Fluoren-9-yl)methoxy)carbonyl)amino)-4-cyclohexylbutanoic acid (9, 500 mg, 1.23 mmol) and HATU (560 mg, 1.47 mmol) were dissolved in DMF (dry, 10 mL), and DIPEA (190 mg, 1.47 mmol) was added. The reaction mixture was stirred for 1.5 h at RT, and rac-1-(methylsulfonyl)piperidin-3-amine (37a, 262 mg, 1.47 mmol) dissolved in DMF (dry, 5 mL) was added. The mixture was stirred for additional 20 h at RT. H₂O (50 mL) was added, and the resulting precipitate was filtered off, washed with H₂O, and dried in a vacuum oven to afford 655 mg (1.15 mmol, 94%) of compound 10d as a colorless solid. TLC: R_f = 0.44 (SiO₂, cyclohexane/EtOAc, 1:3). ¹H NMR (500 MHz, DMSO-d₆, 300 K): (mixture of diastereomeres): δ 7.94 (d, ³J = 7.4 Hz, 1H), 7.89 (d, ³J = 7.5 Hz, 2H), 7.76–7.70 (m, 2H), 7.46 (dd, ³J = 8.3, 3.0 Hz, 1H), 7.42 (t, ³J = 7.5 Hz, 2H), 7.32 (td, ³J =

7.5, 3.4 Hz, 2H), 4.31–4.17 (m, 3H), 3.99–3.90 (m, 1H), 3.75–3.67 (m, 1H), 3.51–3.39 (m, 1H), 3.37–3.27 (m, 1H), 2.87–2.82 (m, 3H), 2.81–2.73 (m, 1H), 2.56–2.49 (m, 1H), 1.83–1.69 (m, 2H), 1.68–1.45 (m, 8H), 1.42–1.27 (m, 1H), 1.26–1.05 (m, 6H), 0.90–0.77 (m, 2H) ppm. ¹³C NMR (126 MHz, DMSO-d₆, 300 K): (mixture of diastereomeres): δ 170.7 (2×), 155.9 (2×), 143.9, 143.8, 140.7 (2×), 127.6, 127.0, 125.4, 120.1, 65.6, 54.8, 54.7, 49.6, 49.4, 46.7, 45.4, 45.1, 45.0, 36.8 (2×), 34.6, 34.2, 33.1, 32.9 (2×), 32.7, 32.6, 29.7, 28.9, 28.6, 26.2, 25.8 (2×), 23.0, 22.9 ppm. MS (ESI pos.) m/z (%): 586.32 (100) ([M + H]⁺ calcd, 586.29), 346.21 (62) ([M − Fmoc + H]⁺ calcd, 346.22).

(9H-Fluoren-9-yl)methyl ((S)-4-Cyclohexyl-1-(((SR)-1-(oxetan-3-yl)piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamate (10e). (S)-2-((((9H-Fluoren-9-yl)methoxy)carbonyl)amino)-4-cyclohexylbutanoic acid (9, 463 mg, 1.14 mmol) and HATU (432 mg, 1.36 mmol) were dissolved in DMF (dry, 4 mL), and DIPEA (147 mg, 1.36 mmol) was added. The reaction mixture was stirred for 1.5 h at RT, and rac-1-(oxetan-3-yl)piperidin-3-amine (37e, 213 mg, 1.36 mmol) dissolved in DMF (dry, 4 mL) was added. The mixture was stirred for additional 4 h at RT. DCM (50 mL) was added, and the organic phase was washed four times with H₂O (50 mL) and dried over MgSO₄. The solvent was evaporated under reduced pressure, and the crude product was purified by column chromatography on silica (cyclohexane/EtOAc 1:1 → 1:20 → EtOAc). Compound 10e was obtained in a yield of 595 mg (1.09 mmol, 96%), as a colorless solid. TLC: R_f = 0.20 (SiO₂, cyclohexane/EtOAc, 1:20). ¹H NMR (500 MHz, DMSO-d₆, 300 K): (mixture of diastereomeres): δ 7.89 (d, ³J = 7.6 Hz, 2H), 7.78–7.69 (m, 3H), 7.47–7.38 (m, 3H), 7.31 (t, ³J = 7.4 Hz, 2H), 4.52–4.43 (m, 2H), 4.41–4.31 (m, 2H), 4.30–4.17 (m, 3H), 3.96–3.87 (m, 1H), 3.76–3.66 (m, 1H), 3.40–3.28 (m, 3H), 2.55–2.32 (m, 2H), 1.92–1.39 (m, 11H), 1.31–1.04 (m, 6H), 0.91–0.76 (m, 2H) ppm. ¹³C NMR (126 MHz, DMSO-d₆, 300 K): (mixture of diastereomeres): δ 171.3, 171.2, 155.9 (2×), 143.9, 143.8, 140.7, 127.6, 127.0, 125.3 (2×), 120.1, 74.6 (3x), 74.5, 65.6 (2×), 58.4 (2×), 54.8, 54.4, 54.3, 49.2, 46.7, 45.1, 38.3, 36.8, 36.7, 33.0, 32.9, 32.7 (2×), 29.7, 29.6, 29.5, 26.2, 25.8, 22.9 ppm. MS (ESI pos.) m/z (%): 546.31 (100) ([M + H]⁺ calcd, 546.34).

(9H-Fluoren-9-yl)methyl ((S)-4-Cyclohexyl-1-(((S)-1-(methylsulfonyl)piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamate (10f). The synthesis of (9H-fluoren-9-yl)methyl ((S)-4-cyclohexyl-1-(((R)-1-(methylsulfonyl)piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamate (10f) was performed analogously to that of compound 10d. (S)-1-(Methylsulfonyl)piperidin-3-amine (37b, 262 mg, 1.47 mmol) was used as a starting material to afford 663 mg (1.17 mmol, 95%) of compound 10f as a colorless solid. TLC: R_f = 0.76 (SiO₂, EtOAc/MeoH, 9:1). ¹H NMR (500 MHz, DMSO-d₆, 300 K): δ 7.91–7.78 (m, 4H), 7.76–7.70 (m, 2H), 7.46–7.59 (m, 1H), 7.41 (td, ³J = 7.4, 1.2 Hz, 2H), 7.34 (td, ³J = 7.4, 1.2 Hz, 1H), 6.66 (d, ³J = 8.5 Hz, 1H), 4.32–4.18 (m, 3H), 3.92–3.84 (m, 1H), 3.78–3.66 (m, 1H), 3.56–3.22 (m, 2H), 2.84 (s, 3H), 2.85–2.74 (m, 1H), 2.67–2.49 (m, 1H), 1.86–1.31 (m, 11H), 1.28–1.04 (m, 6H), 0.90–0.74 (m, 2H) ppm. ¹³C NMR (126 MHz, DMSO-d₆, 300 K): δ 174.6, 172.1, 156.8 (2×), 143.9, 143.8, 140.7 (2×), 128.9, 127.7, 127.3, 123.9, 121.4, 120.0, 61.8, 54.6, 54.5, 49.5, 49.4, 45.4 (2×), 45.0, 44.6, 37.0, 36.8, 34.3, 34.2, 32.9 (2×), 32.8, 32.7, 32.6, 32.4, 29.8, 28.9, 28.7, 26.2 (2×), 25.7, 22.9, 22.7 ppm. MS (ESI pos.) m/z (%): 590.34 (100) ([M + Na]⁺ calcd, 590.27).

(9H-Fluoren-9-yl)methyl ((S)-4-Cyclohexyl-1-(((R)-1-(methylsulfonyl)piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamate (10g). The synthesis of (9H-fluoren-9-yl)methyl ((S)-4-cyclohexyl-1-(((S)-1-(methylsulfonyl)piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamate (10g) was performed analogously to that of compound 10d. (R)-1-(Methylsulfonyl)piperidin-3-amine (37c, 262 mg, 1.47 mmol) was used as a starting material to afford 655 mg (1.15 mmol, 94%) of compound 10g as a light-brownish solid. TLC: R_f = 0.76 (SiO₂, EtOAc/MeOH, 9:1). ¹H NMR (500 MHz, DMSO-d₆, 300 K): δ 7.92–7.73 (m, 4H), 7.46–7.59 (m, 1H, NHCOO), 7.42–7.31 (m, 4H), 6.65 (d, ³J = 8.5 Hz, 1H), 4.33–4.22 (m, 3H), 3.89–3.82 (m, 1H), 3.78–3.66 (m, 1H), 3.53–3.25 (m, 2H), 2.85 (s, 3H), 2.89–2.71 (m, 1H), 2.64–2.47 (m, 1H), 1.82–1.69 (m, 2H), 1.68–1.28 (m,9H), 1.26–1.02 (m, 6H), 0.90–0.74 (m, 2H) ppm. ¹³C NMR (126 MHz, DMSO-d₆, 300 K): δ 174.6, 172.1, 156.8, 143.9, 140.6 140.4, 128.9, 127.7, 127.3, 123.9,

121.4, 120.0, 61.8, 54.7 (2×), 49.6, 49.5, 45.4, 45.3, 45.1, 44.6, 36.9, 36.8, 34.4, 34.3, 32.9 (2×), 32.8, 32.7 (2×), 32.3, 29.8, 28.9, 28.8, 26.1 (2×), 25.8, 23.0, 22.9 ppm. MS (ESI pos.) m/z (%): 590.34 (100) ([M + H]$^+$ calcd, 590.27), 346.22 (38) ([M − Fmoc + H]$^+$ calcd, 346.22).

(S)-2-Amino-4-cyclohexyl-N-ethylbutanamide (11a) was prepared as previously published.[24]

(S)-2-Amino-4-cyclohexyl-N-propylbutanamide (11b). The synthesis of (S)-2-amino-4-cyclohexyl-N-propylbutanamide (11b) was performed according to the general procedure for Fmoc deprotection. The starting material was (9H-fluoren-9-yl)methyl ((S)-4-cyclohexyl-1-(((SR)-1-(cyclopropanecarbonyl)piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamate (10a, 405 mg, 0.716 mmol). The amine was freshly prepared prior to the synthesis of 11b.

(S)-2-Amino-4-cyclohexyl-N-((SR)-1-(cyclopropanecarbonyl)-piperidin-3-yl)butanamide (11c). The synthesis of (S)-2-amino-4-cyclohexyl-N-((SR)-1-(cyclopropanecarbonyl)piperidin-3-yl)-butanamide (11c) was performed according to the general procedure for Fmoc deprotection. The starting material was (9H-fluoren-9-yl)methyl ((S)-4-cyclohexyl-1-(((SR)-1-(cyclopropanecarbonyl)-piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamate (10c, 405 mg, 0.716 mmol). The amine was freshly prepared prior to the synthesis of 11c.

(S)-2-Amino-4-cyclohexyl-N-((SR)-1-(methylsulfonyl)piperidin-3-yl)butanamide (11d). The synthesis of (S)-2-amino-4-cyclohexyl-N-((SR)-1-(methylsulfonyl)piperidin-3-yl)butanamide (11d) was performed according to the general procedure for Fmoc deprotection. The starting material was (9H-fluoren-9-yl)methyl ((S)-4-cyclohexyl-1-(((SR)-1-(methylsulfonyl)piperidin-3-yl)amino)-1-oxobutan-2-yl)-carbamate (10d, 405 mg, 0.714 mmol). The amine was freshly prepared prior to the synthesis of 11d.

(S)-2-Amino-4-cyclohexyl-N-((RS)-1-(oxetan-3-yl)piperidin-3-yl)butanamide (11e). The synthesis of (S)-2-amino-4-cyclohexyl-N-((RS)-1-(oxetan-3-yl)piperidin-3-yl)butanamide (11e) was performed according to the general procedure for Fmoc deprotections. The starting material was (9H-fluoren-9-yl)methyl ((S)-4-cyclohexyl-1-(((S)-1-(methylsulfonyl)piperidin-3-yl)amino)-1-oxobutan-2-yl)-carbamate (10e, 453 mg, 0.830 mmol). The amine was freshly prepared prior to the synthesis of 11e.

(S)-2-Amino-4-cyclohexyl-N-((S)-1-(methylsulfonyl)piperidin-3-yl)butanamide (11f). The synthesis of (S)-2-amino-4-cyclohexyl-N-((S)-1-(methylsulfonyl)piperidin-3-yl) (11f) was performed according to the general procedure for Fmoc deprotection. The starting material was (9H-fluoren-9-yl)methyl ((S)-4-cyclohexyl-1-(((S)-1-(methylsulfonyl)piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamate (10f, 406 mg, 0.716 mmol). The amine was freshly prepared prior to the synthesis of 11f.

(S)-2-Amino-4-cyclohexyl-N-((R)-1-(methylsulfonyl)piperidin-3-yl)butanamide (11g). The synthesis of (S)-2-amino-4-cyclohexyl-N-((R)-1-(methylsulfonyl)piperidin-3-yl) (11g) was performed according to the general procedure for Fmoc deprotection. The starting material was (9H-fluoren-9-yl)methyl ((S)-4-cyclohexyl-1-(((R)-1-(methylsulfonyl)piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamate (10g, 835 mg, 1.47 mmol). The amine was freshly prepared prior to the synthesis of 11g.

(S)-5-Amino-N-(4-((4-cyclohexyl-1-oxo-1-(propylamino)butan-2-yl)carbamoyl)benzyl)-1-phenyl-1H-pyrazole-4-carboxamide (12). 4-((5-Amino-1-phenyl-1H-pyrazole-4-carboxamido)methyl)benzoic acid (12, 100 mg, 0.30 mmol) and HATU (152 mg, 0.40 mmol) were dissolved in DMF (5 mL), and DIPEA (58 mg, 0.45 mmol) was added. The solution was stirred for 1.5 h at RT, and (S)-2-amino-4-cyclohexyl-N-propylbutanamide (11b, 81 mg, 0.36 mmol), diluted in DMF (5 mL) and DIPEA (58 mg, 0.45 mmol), was added. The mixture was stirred for 20 h at RT. DCM (100 mL) was added, and the organic phase was washed five times with H$_2$O (20 mL) and dried over MgSO$_4$. The solvent was evaporated under reduced pressure, and the crude product was purified using column chromatography on silica (DCM/MeOH 20:1). Compound 12 was obtained in a yield of 121 mg (0.22 mmol, 75%), as a yellowish solid. TLC: R_f = 0.17 (SiO$_2$, DCM/MeOH 20:1). HPLC: t_R = 9.31 min, purity: >95%. ^{1}H NMR (500 MHz, DMSO-d_6, 300 K): δ 8.55−8.47 (m, 1H), 8.28 (d, ^3J = 8.1 Hz, 1H), 7.98 (s, 1H), 7.92 (t, ^3J = 5.7 Hz, 1H), 7.85 (d, ^3J = 8.3 Hz, 2H), 7.59−7.49 (m, 4H),

7.41−7.31 (m, 3H), 6.38 (br s, 2H), 4.49−4.43 (m, 2H), 4.39−4.32 (m, 1H), 3.10−2.93 (m, 2H), 1.81−1.71 (m, 1H), 1.70−1.55 (m, 6H), 1.40 (sext, ^3J = 7.3 Hz, 2H), 1.28−1.04 (m, 6H), 0.90−0.77 (m, 5H) ppm. ^{13}C NMR (126 MHz, DMSO-d_6, 300 K): δ 171.7, 166.0, 164.2, 149.2, 143.6, 141.4, 138.4, 138.2, 132.7, 129.4, 127.6, 127.1, 127.0, 126.8 (2×), 123.1, 97.4, 53.6, 41.3, 36.8, 33.3, 32.9, 32.7, 29.4, 26.2, 25.8 (2×), 22.3, 11.4 ppm. MS (ESI pos.) m/z (%): 545.32 (100) ([M + H]$^+$ calcd, 545.33), 486.27 (30) ([M$_{fc}$]$^+$ calcd, 486.25). HRMS (FTMS + p MALDI) m/z: 545.3222 [M + H]$^+$ calcd for [C$_{31}$H$_{41}$N$_6$O$_3$]$^+$, 545.3240.

5-Amino-N-(4-(((S)-4-cyclohexyl-1-(((SR)-1-(methylsulfonyl)-piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamoyl)benzyl)-1-phe-nyl-1H-pyrazole-4-carboxamide (13). 4-((5-Amino-1-phenyl-1H-pyrazole-4-carboxamido)methyl)benzoic acid (8, 200 mg, 0.59 mmol) and HATU (280 mg, 0.74 mmol) were dissolved in DMF (5 mL), and DIPEA (115 mg, 0.89 mmol) was added. The solution was stirred for 1.5 h at RT, and (S)-2-amino-4-cyclohexyl-N-((SR)-1-(methylsulfonyl)piperidin-3-yl)butanamide (11d, 247 mg, 0.71 mmol), diluted in DMF (5 mL) and DIPEA (115 mg, 0.89 mmol), was added. The mixture was stirred for 20 h at RT. DCM (50 mL) was added, and the organic phase was washed five times with H$_2$O (20 mL) and dried over MgSO$_4$. The solvent was evaporated under reduced pressure, and the crude product was triturated with cyclohexane/EtOAc (1:1) to afford 78 mg (0.11 mmol, 19%) of compound 13 as a colorless solid. TLC: R_f = 0.14 (SiO$_2$, cyclohexane/EtOAc 1:3). HPLC: (mixture of diastereomers) $t_{R,1}$ = 8.90, $t_{R,2}$ = 9.08 min, purity: >95%. ^{1}H NMR (500 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 8.52 (t, ^3J = 6.0 Hz, 1H), 8.32 (d, ^3J = 8.1 Hz, 1H), 8.03 (d, ^3J = 7.5 Hz, 1H), 7.99 (s, 1H), 7.85 (d, ^3J = 8.4 Hz, 2H), 7.60−7.49 (m, 4H), 7.42−7.34 (m, 3H), 6.38 (br s, 2H), 4.47 (d, ^3J = 6.0 Hz, 2H), 4.44−4.35 (m, 1H), 3.78−3.67 (m, 1H), 3.53−3.40 (m, 1H), 3.37−3.28 (m, 1H), 2.87−2.82 (m, 3H), 2.81−2.71 (m, 1H), 2.58−2.51 (m, 1H), 1.85−1.46 (m, 10H), 1.45−1.30 (m, 1H), 1.29−1.04 (m, 6H), 0.90−0.77 (m, 2H) ppm. ^{13}C NMR (126 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 171.7, 166.1 (2×), 164.1, 149.2, 143.6, 138.4, 138.2, 132.7, 129.4, 127.6, 127.1, 126.8, 123.1, 97.4, 53.6, 53.5, 49.6, 49.4, 45.4, 45.2, 45.1, 41.3, 36.8 (2×), 34.3 (2×), 33.3, 32.9 (2×), 32.7 (2×), 29.5, 29.0, 28.7, 26.2, 25.8 (2×), 23.0 (2×) ppm. MS (ESI neg.) m/z (%): 662.27 (100) ([M − H]$^-$ calcd, 662.31), 698.30 (50) ([M + Cl]$^-$ calcd, 698.77). HRMS (FTMS + p MALDI) m/z: 702.2830 [M + K]$^+$ calcd for [C$_{34}$H$_{45}$KN$_7$O$_5$S]$^+$, 702.2840.

5-Amino-N-(4-(((S)-4-cyclohexyl-1-(((SR)-1-(cyclopropanecarbonyl)piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamoyl)benzyl)-1-phenyl-1H-pyrazole-4-carboxamide (14). 4-((5-Amino-1-phenyl-1H-pyrazole-4-carboxamido)methyl)benzoic acid (8, 200 mg, 0.59 mmol) and HATU (280 mg, 0.74 mmol) were dissolved in DMF (5 mL), and DIPEA (115 mg, 0.89 mmol) was added. The solution was stirred for 1.5 h at RT, and (S)-2-amino-4-cyclohexyl-N-((SR)-1-(cyclopropanecarbonyl)piperidin-3-yl)butanamide (11c, 240 mg, 0.71 mmol), diluted in DMF (5 mL) and DIPEA (115 mg, 0.89 mmol), was added. The mixture was stirred for 20 h at RT. DCM (50 mL) was added, and the organic phase was washed five times with H$_2$O (20 mL) and dried over MgSO$_4$. The solvent was evaporated under reduced pressure, and the crude product was triturated with cyclohexane/EtOAc (1:1) to afford 120 mg (0.18 mmol, 31%) of compound 14 as a colorless solid. TLC: R_f = 0.21 (SiO$_2$, EtOAc/MeOH 10:1). HPLC: (mixture of diastereomers) $t_{R,1}$ = 8.96, $t_{R,2}$ = 9.08 min, purity: 91%. ^{1}H NMR (500 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 8.52 (t, ^3J = 6.1 Hz, 1H), 8.30 (d, ^3J = 8.1 Hz, 1H), 8.07−7.88 (m, 2H), 7.85 (d, ^3J = 8.3 Hz, 2H), 7.60−7.49 (m, 4H), 7.43−7.34 (m, 3H), 6.38 (br s, 2H), 4.53−4.31 (m, 3H), 4.20−3.46 (m, 3H), 3.27−3.01 (m, 1H), 2.76−2.56 (m, 1H), 2.54−2.45 (m, 1H), 2.02−1.31 (m, 12H), 1.30−1.02 (m, 6H), 0.92−0.77 (m, 2H), 0.92−0.77 (m, 4H) ppm. ^{13}C NMR (126 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 172.0, 171.4, 171.0, 170.9, 166.1, 164.1, 149.2, 143.6, 138.4, 138.2, 132.7, 132.6, 129.4, 127.6, 127.1, 126.8, 123.1, 97.4, 53.4, 49.0, 46.1, 45.9, 45.1, 41.8, 41.3, 36.9, 33.5, 33.3, 33.0, 32.7, 29.9, 29.7, 29.4, 29.3, 26.2, 25.8 (2×), 24.1, 22.5, 10.4, 10.3, 7.0, 6.8 ppm. MS (ESI neg.) (%): 652.31 (100) ([M − H]$^-$ calcd, 652.36). HRMS (FTMS + p MALDI) m/z: 692.3311 [M + K]$^+$ calcd for [C$_{37}$H$_{47}$KN$_7$O$_4$]$^+$, 692.3321.

5-Amino-N-(4-(((S)-4-cyclohexyl-1-(((SR)-1-(oxetan-3-yl)-piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamoyl)benzyl)-1-phenyl-1H-pyrazole-4-carboxamide (15). 4-((5-Amino-1-phenyl-1H-pyrazole-4-carboxamido)methyl)benzoic acid (8, 233 mg, 0.69 mmol) and HATU (316 mg, 0.83 mmol) were dissolved in DMF (5 mL), and DIPEA (134 mg, 1.038 mmol) was added. The solution was stirred for 1 h at RT, and (S)-2-amino-4-cyclohexyl-N-((SR)-1-(oxetan-3-yl)piperidin-3-yl)butanamide (11e, 268 mg, 0.83 mmol), diluted in DMF (5 mL), was added. The mixture was stirred for 20 h at RT. DCM (50 mL) was added, and the organic phase was washed five times with H_2O (20 mL) and dried over $MgSO_4$. The solvent was evaporated under reduced pressure, and the crude product was purified using column chromatography on silica (cyclohexane/EtOAc 1:20 → EtOAc → EtOH) to afford 175 mg (0.273 mmol, 39%) of compound 15 as a colorless solid. TLC: R_f = 0.16 (SiO_2, EtOAc/MeOH 10:1). HPLC: (mixture of diastereomers) $t_{R,1}$ = 7.57, $t_{R,2}$ = 7.61 min, purity: >95%. ¹H NMR (500 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 8.60 (t, ³J = 5.9 Hz, 1H), 8.40 (d, ³J = 8.2 Hz, 1H), 8.02 (s, 1H), 7.91 (d, ³J = 7.9 Hz, 1H), 7.85 (dd, ³J = 8.5, 1.9 Hz, 2H), 7.59−7.49 (m, 4H), 7.41−7.35 (m, 3H), 6.38 (br s, 2H), 4.53−4.44 (m, 4H), 4.43−4.31 (m, 3H), 3.76−3.67 (m, 1H), 3.48−3.26 (m, 3H), 2.55−2.46 (m, 1H), 2.45−2.35 (m, 1H), 1.89−1.54 (m, 10H), 1.51−1.38 (m, 1H), 1.31−1.04 (m, 6H), 0.89−0.77 (m, 2H) ppm. ¹³C NMR (126 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 171.3 (2×), 166.1 (2×), 164.2, 149.2, 143.7, 143.6, 138.5, 138.3, 132.7 (2×), 129.4, 127.6, 127.1, 126.8, 123.1, 97.5, 74.7, 74.6, 58.4, 54.4 (2×), 53.7, 53.6, 49.2, 45.2 (2×), 41.3, 36.8 (2×), 33.3, 33.0, 32.7, 29.7, 29.5, 29.4, 26.2, 25.8 (2×), 23.0, 29.9 ppm. MS (ESI pos.) m/z (%): 642.51 (100) ([M − H]⁺ calcd, 642.38). HRMS (FTMS + p MALDI) m/z: 642.3760 [M + H]⁺ calcd for [$C_{36}H_{48}N_7O_4$]⁺, 642.3762.

Methyl 4-((3-Phenyl-1H-pyrazole-5-carboxamido)methyl)-benzoate (18a). 3-Phenyl-1H-pyrazole-5-carboxylic acid (16, 500 mg, 2.66 mmol), HOBt (431 mg, 3.19 mmol), and EDC-HCl (611 mg, 3.19 mmol) were dissolved in DMF (dry, 10 mL), and DIPEA (512 mg, 3.99 mmol) was added. The mixture was stirred for 1 h at RT, and a solution of methyl 4-(aminomethyl)benzoate hydrochloride (17a, 688 mg, 3.19 mmol) dissolved in DMF (dry, 15 mL) and DIPEA (512 mg, 3.99 mmol) was added. The reaction mixture was stirred for additional 20 h at RT. DCM (150 mL) was added, and the organic phase was washed four times with H_2O (150 mL). The organic phase was separated and dried over $MgSO_4$, and the solvent was removed under reduced pressure. The crude product was dissolved in DCM and precipitated with cyclohexane to afford 711 mg (2.12 mmol, 80%) of compound 18a, as a colorless solid. TLC: R_f = 0.70 (SiO_2, cyclohexane/EtOAc 1:3). ¹H NMR (600 MHz, DMSO-d_6, 300 K): δ 13.70 (s, 1H), 9.22−8.73 (s, 1H), 7.93 (d, ³J = 8.1 Hz, 2H), 7.80 (d, ³J = 7.3 Hz, 2H), 7.51−7.42 (m, 4H), 7.39−7.32 (m, 1H), 4.60−4.48 (m, 2H), 3.84 (s, 3H) ppm. ¹³C NMR (150 MHz, DMSO-d_6, 300 K): δ 166.1, 161.8, 147.7, 145.4, 140.8, 129.2 (2×), 128.9, 128.1, 127.3, 125.2, 102.5, 52.0, 41.8 ppm. MS (ESI neg.) m/z (%): 334.07 (100) ([M − H]⁻ calcd, 334.12).

Methyl 4-(1-(3-Phenyl-1H-pyrazole-5-carboxamido)ethyl)-benzoate (18b). 3-Phenyl-1H-pyrazole-5-carboxylic acid (16, 1.00 g, 5.314 mmol), HOBt (862 mg, 6.38 mmol), and EDC-HCl (1.22 g, 6.38 mmol) were dissolved in DMF (dry, 10 mL), and DIPEA (1.03 g, 7.97 mmol) was added. The mixture was stirred for 1.5 h at RT, and a solution of methyl 4-(1-aminoethyl)benzoate hydrochloride (17b, 1.38 g, 6.38 mmol) dissolved in DMF (dry, 10 mL) and DIPEA (1.03 g, 7.97 mmol) was added. The reaction mixture was stirred for additional 20 h at RT. DCM (100 mL) and H_2O (200 mL) were added, and the aqueous phase was extracted two times with DCM (50 mL). The organic phase was separated and dried over $MgSO_4$, and the solvent was removed under reduced pressure. The crude product was purified using column chromatography on silica (cyclohexane/EtOAc 1:1 → 1:3). Compound 18b was obtained in a yield of 1.606 g (4.80 mmol, 90%) as a colorless solid. TLC: R_f = 0.29 (SiO_2, cyclohexane/EtOAc 1:1). ¹H NMR (600 MHz, DMSO-d_6, 300 K): δ 7.94 (d, ³J = 8.5 Hz, 2H), 7.58 (d, ³J = 6.9 Hz, 2H), 7.41 (d, ³J = 8.3 Hz, 2H), 7.38−7.30 (m, 4H), 7.06 (s, 1H), 5.35 (p, ³J = 7.4 Hz, 1H), 3.88 (s, 3H), 1.56 (d, ³J = 7.0 Hz, 3H) ppm. ¹³C NMR (126 MHz, DMSO-d_6, 300 K): δ 167.1, 161.4, 148.6,

146.2, 130.1, 129.2 (2×), 128.9, 126.3, 126.2, 125.7, 103.4, 52.3, 48.7, 22.1 ppm. MS (ESI neg.) m/z (%): 348.11 (100) ([M − H]⁻ calcd, 348.13), 697.20 (24) ([2M − H]⁻ calcd, 697.27).

4-((3-Phenyl-1H-pyrazole-5-carboxamido)methyl)benzoic Acid (19a). Methyl 4-((3-phenyl-1H-pyrazole-5-carboxamido)methyl)-benzoate (18a, 600 mg, 1.79 mmol) was dissolved in a mixture of THF/H_2O (10 mL, 1:1), and LiOH × H_2O (214 mg, 8.95 mmol) was added. The suspension was stirred for 24 h at RT and then acidified with HCl (aq, 10 mL, 1 M). The precipitate was filtered, washed with H_2O, and dried under reduced pressure in a vacuum oven to afford 544 mg (1.69 mmol, 95%) of compound 19a as a colorless solid. TLC: R_f = 0.30 (SiO_2, cyclohexane/EtOAc 1:3). ¹H NMR (600 MHz, DMSO-d_6, 300 K): δ 13.70−12.68 (m, 2H), 8.95 (s, 1H), 7.91 (d, ³J = 8.0 Hz, 2H), 7.80 (d, ³J = 7.6 Hz, 2H), 7.50−7.41 (m, 4H), 7.36 (t, ³J = 7.6 Hz, 1H), 7.17 (s, 1H), 4.54 (d, ³J = 6.2 Hz, 2H) ppm. ¹³C NMR (150 MHz, DMSO-d_6, 300 K): δ 167.2, 157.0, 147.7, 144.8, 140.8, 129.4, 129.3, 129.0, 128.2, 127.2 (2×), 125.2, 102.5, 41.8 ppm. MS (ESI neg.) m/z (%): 320.10 (100) ([M − H]⁻ calcd, 320.10), 356.08 (56) ([M + Cl]⁻ calcd, 356.56).

4-(1-(3-Phenyl-1H-pyrazole-5-carboxamido)ethyl)benzoic Acid (19b). Methyl 4-(1-(3-phenyl-1H-pyrazole-5-carboxamido)ethyl)-benzoate (18b, 200 mg, 0.57 mmol) was dissolved in a mixture of THF/H_2O (10 mL, 1:1), and LiOH × H_2O (69 mg, 2.86 mmol) was added. The suspension was stirred for 24 h at RT and then acidified with HCl (aq, 10 mL, 1 M). The precipitate was filtered, washed with H_2O, and dried under reduced pressure in a vacuum oven to yield 178 mg (0.531 mmol, 93%) of compound 19b as a colorless solid. TLC: R_f = 0.33 (SiO_2, cyclohexane/EtOAc 1:3). ¹H NMR (600 MHz, DMSO-d_6, 300 K): δ 13.95−12.45 (m, 2H), 8.72 (s, 1H), 7.91 (d, ³J = 8.2 Hz, 2H), 7.80 (d, ³J = 7.6 Hz, 2H), 7.52 (d, ³J = 8.2 Hz, 2H), 7.46 (t, ³J = 7.6 Hz, 2H), 7.35 (t, ³J = 7.5 Hz, 1H), 7.19 (s, 1H), 5.21 (p, ³J = 7.1 Hz, 1H), 1.56 (d, ³J = 7.1 Hz, 3H) ppm. ¹³C NMR (150 MHz, DMSO-d_6, 300 K): δ 168.1, 150.7, 148.6, 146.2, 130.4, 103.2, 129.9, 129.1, 127.3, 126.2, 125.2, 103.6, 47.8, 21.8 ppm. MS (ESI neg.) m/z (%): 334.10 (100) ([M − H]⁻ calcd, 334.12), 370.05 (22) [M + Cl]⁻ calcd, 370.58).

N-(4-(((S)-4-Cyclohexyl-1-(((SR)-1-(methylsulfonyl)piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamoyl)benzyl)-3-phenyl-1H-pyra-zole-5-carboxamide (20). 4-((3-Phenyl-1H-pyrazole-5-carboxamido)methyl)benzoic acid (19a, 200 mg, 0.62 mmol) and HATU (284 mg, 0.75 mmol) were dissolved in DMF (5 mL), and DIPEA (96 mg, 0.75 mmol) was added. The solution was stirred for 1.5 h at RT, and (S)-2-amino-4-cyclohexyl-N-((SR)-1-(methylsulfonyl)-piperidin-3-yl)butanamide (11d, 258 mg, 0.75 mmol), diluted in DMF (5 mL), was added. The mixture was stirred for 20 h at RT. DCM (100 mL) was added, and the organic phase was washed four times with H_2O (100 mL) and dried over $MgSO_4$. The solvent was evaporated under reduced pressure, and the crude product was purified using flash column chromatography on silica (cyclohexane/EtOAc 1:1 → 1:3 → 1:20 → EtOAc) to afford 113 mg (0.17 mmol, 28%) of compound 20 as a colorless solid. TLC: R_f = 0.44 (SiO_2, cyclohexane/EtOAc 1:10). HPLC: (mixture of diastereomers) $t_{R,1}$ = 9.03 min; $t_{R,2}$ = 9.20 min, purity: >95%. ¹H NMR (600 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 13.68 (br s, 1H), 8.86 (br s, 1H), 8.32 (d, ³J = 8.0 Hz, 1H), 8.01 (d, ³J = 7.6 Hz, 1H), 7.91−7.71 (m, 4H), 7.54−7.29 (m, 5H), 7.12 (br s, 1H), 4.60−4.32 (m, 3H), 3.82−3.64 (m, 1H), 3.55−3.41 (m, 1H), 3.37−3.24 (m, 1H), 2.85 (s, 3H), 2.82−2.70 (m, 1H), 2.60−2.53 (m, 1H), 1.85−1.46 (m, 10H), 1.44−1.31 (m, 1H), 1.26−1.02 (m, 6H), 0.90−0.76 (m, 2H) ppm. ¹³C NMR (150 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 171.6, 166.0, 162.1, 149.5, 143.9, 132.7, 128.9, 127.5, 126.9, 125.2, 102.9, 53.6, 53.4, 49.5, 49.4, 45.3, 45.3, 45.2, 45.1, 41.8, 36.8, 34.4, 34.3, 33.2, 32.9, 32.6, 29.4, 28.9, 28.6, 26.1, 25.7, 22.9 ppm. MS (ESI neg.) m/z (%): 647.22 (100) ([M − H]⁻ calcd, 647.30). HRMS (FTMS + p MALDI) m/z: 671.2980 [M + Na]⁺ calcd for [$C_{34}H_{44}N_6NaO_4S$]⁺, 671.2992.

N-((SR)-1-(4-(((S)-4-Cyclohexyl-1-(((SR)-1-(methylsulfonyl)-piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamoyl)phenyl)ethyl)-3-phenyl-1H-pyrazole-5-carboxamide Carboxamide (21). 4-(1-(3-Phenyl-1H-pyrazole-5-carboxamido)ethyl)benzoic acid (19b, 200 mg, 0.60 mmol) and HATU (309 mg, 0.72 mmol) were dissolved in

DMF (5 mL), and DIPEA (105 mg, 0.72 mmol) was added. The solution was stirred for 1.5 h at RT, and (S)-2-amino-4-cyclohexyl-N-((SR)-1-(methylsulfonyl)piperidin-3-yl)butanamide (11d, 281 mg, 0.72 mmol), diluted in DMF (5 mL), was added. The mixture was stirred for 20 h at RT. DCM (100 mL) was added, and the organic phase was washed four times with H$_2$O (100 mL) and dried over MgSO$_4$. The solvent was evaporated under reduced pressure, and the crude product was purified using flash column chromatography on silica (cyclohexane/EtOAc 1:3 → EtOAc) to afford 89 mg (0.13 mmol, 30%) of compound 21 as a colorless solid. TLC: R_f = 0.70 (SiO$_2$, EtOAc/MeOH 1:10). HPLC: (mixture of diastereomers) $t_{R,1}$ = 9.21 min; $t_{R,2}$ = 9.34 min, purity: >95%. ^{1}H NMR (400 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 13.64 (br s, 1H), 8.97–8.76 (m, 0.4H), 8.66–8.45 (m, 0.6H), 8.30 (d, ^3J = 7.6 Hz, 1H), 8.04–7.93 (m, 1H), 7.89–7.82 (m, 2H), 7.80 (d, ^3J = 7.2 Hz, 2H), 7.54–7.29 (m, 5.4H), 7.06 (s, 0.6H), 5.20 (p, ^3J = 7.6 Hz, 1H), 4.45–4.33 (m, 1H), 3.81–3.66 (m, 1H), 3.54–3.40 (m, 1H), 3.39–3.27 (m, 1H), 2.85 (s, 3H), 2.81–2.70 (m, 1H), 2.59–2.51 (m, 1H), 1.85–1.46 (m, 13H), 1.44–1.31 (m, 1H), 1.29–1.05 (m, 6H), 0.92–0.72 (m, 2H) ppm. ^{13}C NMR (101 MHz, DMSO-d_6, 300 K): (mixture of diastereomers: δ 171.6, 166.2, 161.9, 147.7, 144.9, 132.7, 129.1, 128.8, 127.9, 127.7, 127.6, 126.0, 125.4, 102.8, 53.5, 53.4, 49.6, 49.4, 45.4, 45.2, 45.1, 36.8, 34.4, 33.3 (2×), 32.9 (2×), 32.7, 29.4, 28.9, 28.7, 26.1, 25.8 (2×), 23.0 (2×), 22.0 (2×) ppm. MS (ESI neg.) m/z (%): 707.29 (93) ([M + EtOH − H]$^-$ calcd, 707.36), 661.29 (100) ([M − H]$^-$ calcd, 661.32), 418.17 (30) ([M$_6$ + EtOH − H]$^-$ calcd, 418.23), 372.19 (42) ([M$_6$]$^-$ calcd, 372.20). HRMS (FTMS + p MALDI) m/z:685.3138 [M + Na]$^+$ calcd for [C$_{35}$H$_{46}$N$_6$NaO$_5$S]$^+$, 685.3148.

N-((SR)-1-(4-(((S)-4-Cyclohexyl-1-(((S)-1-(methylsulfonyl)piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamoyl)phenyl)ethyl)-3-phenyl-1H-pyrazole-5-carboxamide (22). 4-(1-(3-Phenyl-1H-pyrazole-5-carboxamido)ethyl)benzoic acid (19b, 200 mg, 0.60 mmol) and HATU (272 mg, 0.72 mmol) were dissolved in DMF (5 mL), and DIPEA (93 mg, 0.72 mmol) was added. The solution was stirred for 1.5 h at RT, and (S)-2-amino-4-cyclohexyl-N-((S)-1-(methylsulfonyl)-piperidin-3-yl)butanamide (11f, 248 mg, 0.72 mmol), diluted in DMF (5 mL), was added. The mixture was stirred for 20 h at RT. DCM (100 mL) was added, and the organic phase was washed four times with H$_2$O (100 mL) and dried over MgSO$_4$. The solvent was evaporated under reduced pressure, and the crude product was purified using flash column chromatography on silica (cyclohexane/EtOAc 1:1 → 1:3 → 1:20) to afford 121 mg (0.18 mmol, 41%) of compound 22 as a colorless solid. TLC: R_f = 0.07 (SiO$_2$, cyclohexane/EtOAc 1:3). HPLC: (mixture of diastereomers) t_R = 9.21 min, purity: >95%. ^{1}H NMR (600 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 13.66–13.57 (m, 1H), 8.88 (d, ^3J = 8.0 Hz, 0.4H), 8.56 (d, ^3J = 8.4 Hz, 0.6H), 8.34–8.27 (m, 1H), 8.03–7.96 (m, 1H), 7.89–7.82 (m, 2H), 7.80 (d, ^3J = 7.6 Hz, 2H), 7.52–7.42 (m, 4H), 7.40–7.30 (m, 1.4H), 7.06 (s, 0.6H), 5.20 (p, ^3J = 7.3 Hz, 1H), 4.39 (q, ^3J = 8.0 Hz, 1H), 3.77–3.67 (m, 1H), 3.52–3.46 (m, 1H), 3.37–3.31 (m, 1H), 2.85 (s, 3H), 2.78–2.71 (m, 1H), 2.56–2.49 (m, 1H), 1.81–1.46 (m, 13H), 1.39–1.30 (m, 1H), 1.28–1.04 (m, 6H), 0.89–0.79 (m, 2H) ppm. ^{13}C NMR (150 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 171.6, 166.1, 161.0, 148.2, 147.8, 143.5, 132.6, 129.1, 128.8, 128.4, 127.5, 125.9, 125.8, 125.4, 102.8, 102.8, 53.5, 49.5, 47.7, 45.3, 45.2, 36.8, 34.4, 33.2, 32.9, 32.7, 29.4, 28.9, 26.1, 25.8, 25.7, 23.0, 21.9 ppm. MS (ESI neg.) m/z (%): 707.35 (71) ([M + EtOH − H]$^-$ calcd, 707.36), 661.26 (100) ([M − H]$^-$ calcd, 661.31), 372.24 (65) ([M$_6$]$^-$ calcd, 372.20). HRMS (FTMS + p MALDI) m/z: 685.3140 [M + Na]$^+$ calcd for [C$_{35}$H$_{46}$N$_6$NaO$_5$S]$^+$, 685.3148.

N-((SR)-1-(4-(((S)-4-Cyclohexyl-1-(((R)-1-(methylsulfonyl)piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamoyl)phenyl)ethyl)-3-phenyl-1H-pyrazole-5-carboxamide (23). 4-(1-(3-Phenyl-1H-pyrazole-5-carboxamido)ethyl)benzoic acid (19b, 200 mg, 0.60 mmol) and HATU (309 mg, 0.72 mmol) were dissolved in DMF (5 mL), and DIPEA (105 mg, 0.72 mmol) was added. The solution was stirred for 1.5 h at RT, and (S)-2-amino-4-cyclohexyl-N-((R)-1-(methylsulfonyl)-piperidin-3-yl)butanamide (11g, 281 mg, 0.72 mmol), diluted in DMF (5 mL), was added. The mixture was stirred for 20 h at RT. DCM (100 mL) was added, and the organic phase was washed four times with H$_2$O

(100 mL) and dried over MgSO$_4$. The solvent was evaporated under reduced pressure, and the crude product was purified using flash column chromatography on silica (cyclohexane/EtOAc 1:1 → 1:3 → 1:20) to afford 102 mg (0.15 mmol, 34%) of compound 23 as a colorless solid. TLC: R_f = 0.63 (SiO$_2$, EtOAc/MeOH 1:10). HPLC: (mixture of diastereomers) t_R = 9.31 min, purity: >95%. ^{1}H NMR (600 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 13.68–13.56 (m, 1H), 8.94–8.82 (m, 0.4H), 8.63–8.51 (m, 0.6H), 8.35–8.26 (m, 1H), 7.99 (d, ^3J = 7.9 Hz, 1H), 7.88–7.81 (m, 2H), 7.79 (d, ^3J = 7.7 Hz, 2H), 7.51–7.29 (m, 5.4H), 7.06 (s, 0.6H), 5.20 (p, ^3J = 7.4 Hz, 1H), 4.44–4.36 (m, 1H), 3.76–3.67 (m, 1H), 3.46–3.41 (m, 1H), 3.34–3.27 (m, 1H), 2.84 (s, 3H), 2.80–2.77 (m, 1H), 2.58–2.51 (m, 1H), 1.84–1.33 (m, 14H), 1.30–1.05 (m, 6H), 0.90–0.77 (m, 2H) ppm. ^{13}C NMR (150 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 171.6, 166.1, 160.7, 148.2, 147.8, 143.4, 132.6, 129.0 (2×), 127.5, 125.9, 125.3 (2×), 125.0, 102.1, 53.4, 49.4, 45.3, 45.1, 45.0, 36.8, 34.3, 33.2, 32.9, 32.6, 29.4, 28.6, 26.1, 25.8, 23.0, 21.9 (2×) ppm. MS (ESI neg.) m/z (%): 707.32 (48) ([M + EtOH − H]$^-$ calcd, 707.36), 661.31 (41) ([M − H]$^-$ calcd, 661.32), 418.19 (57) ([M$_6$ + EtOH − H]$^-$ calcd, 418.23), 372.19 (100) ([M$_6$]$^-$ calcd, 372.20). HRMS (FTMS + p MALDI) m/z: 685.3144 [M + Na]$^+$ calcd for [C$_{35}$H$_{46}$N$_6$NaO$_5$S]$^+$, 685.3148.

(S)-N-(4-((4-Cyclohexyl-1-oxo-1-(propylamino)butan-2-yl)carbamoyl)benzyl)-3-phenyl-1H-pyrazole-5-carboxamide (24). 4-((3-Phenyl-1H-pyrazole-5-carboxamido)methyl)benzoic acid (19a, 200 mg, 0.62 mmol) and HATU (284 mg, 0.75 mmol) were dissolved in DMF (5 mL), and DIPEA (96 mg, 0.75 mmol) was added. The solution was stirred for 1.5 h at RT, and (S)-2-amino-4-cyclohexyl-N-propylbutanamide (11b, 169 mg, 0.75 mmol), diluted in DMF (5 mL), was added. The mixture was stirred for 20 h at RT. DCM (100 mL) was added, and the organic phase was washed four times with H$_2$O (150 mL) and once with a NaCl solution (sat., 50 mL) and dried over MgSO$_4$. The solvent was evaporated under reduced pressure, and the crude product was precipitated from EtOAc to afford 197 mg (0.37 mmol, 60%) of compound 24 as a colorless solid. TLC: R_f = 0.34 (SiO$_2$, cyclohexane/EtOAc 1:3). HPLC: t_R = 9.44 min, purity: >95%. ^{1}H NMR (250 MHz, DMSO-d_6, 300 K): (mixture of rotamers): δ 13.67 (s, 1H), 9.18–9.02 (m, 0.4H), 8.89–8.73 (m, 0.6H), 8.26 (d, ^3J = 7.9 Hz, 1H), 7.98–7.71 (m, 5H), 7.55–7.25 (m, 5.4H), 7.09 (s, 0.6H), 4.60–4.42 (m, 2H), 4.41–4.28 (m, 1H), 3.15–2.90 (m, 2H), 1.80–1.51 (m, 7H), 1.40 (sext., ^3J = 7.3 Hz, 2H), 1.28–1.04 (m, 6H), 0.92–0.72 (m, 5H) ppm. ^{13}C NMR (126 MHz, DMSO-d_6, 300 K): (mixture of rotamers): δ 171.7, 166.0, 161.8, 147.7, 143.5, 143.3, 132.7, 129.1, 128.8, 128.5, 127.6, 126.9, 125.3, 125.1, 102.8, 53.7, 41.8, 36.8, 33.3, 32.9, 32.7, 29.4, 26.2, 25.8 (2×), 22.3, 11.4 ppm. MS (ESI neg.) m/z (%): 528.14 (100) ([M − H]$^-$ calcd, 528.30). HRMS (FTMS + p MALDI) m/z: 552.2922 [M + Na]$^+$ calcd for [C$_{31}$H$_{39}$N$_5$NaO$_3$]$^+$, 552.2945.

Methyl 4-(((3-Phenyl-3H-[1,2,3]triazolo[4,5-d]pyrimidin-7-yl)amino)methyl)benzoate (26a). 7-Chloro-3-phenyl-3H-[1,2,3]-triazolo[4,5-d]pyrimidine (25a, 250 mg, 1.08 mmol) was diluted in THF (dry, 10 mL), and DIPEA (140 mg, 1.08 mmol) was added. In a separate flask, methyl 4-(aminomethyl)benzoate hydrochloride (17a, 218 mg, 1.08 mmol) and DIPEA (140 mg, 1.08 mmol) were dissolved in THF (dry, 10 mL) and stirred at RT for 5 min. This suspension was then added to the first solution, und the mixture was heated to 60 °C for 20 h. EtOAc (150 mL) was added, and the organic phase was washed three times with H$_2$O (50 mL). The organic phase was separated and dried over MgSO$_4$, and the solvent was removed under reduced pressure. The crude product was precipitated from cyclohexane to afford 324 mg (0.899 mmol, 83%) of compound 26a as a colorless solid. TLC: R_f = 0.72 (SiO$_2$, cyclohexane/EtOAc 1:1). ^{1}H NMR (600 MHz, DMSO-d_6, 300 K): (mixture of rotamers): δ 9.73 (t, ^3J = 6.1 Hz, 0.8H), 9.35 (t, ^3J = 7.0 Hz, 0.2H), 8.47 (s, 0.8H), 8.43 (s, 0.2H), 8.15 (d, ^3J = 7.9 Hz, 1.6H), 8.15 (d, ^3J = 7.9 Hz, 0.4H), 7.92 (d, ^3J = 7.9 Hz, 2H), 7.66 (t, ^3J = 7.9 Hz, 2H), 7.58–7.49 (m, 3H), 5.32 (d, ^3J = 7.0 Hz, 0.4H), 4.88 (d, ^3J = 6.1 Hz, 1.6H), 3.83 (s, 3H) ppm. ^{13}C NMR (150 MHz, DMSO-d_6, 300 K): (mixture of rotamers): δ 166.0, 157.4, 157.0, 154.5, 148.1, 144.5, 135.7, 129.6, 129.4, 129.3, 128.5, 128.3, 127.4,

pubs.acs.org/jmc

124.8, 121.5, 121.4, 52.0, 43.0 ppm. MS (ESI neg.) *m/z* (%): 359.05 (100) ([M − H]⁻ calcd, 359.12).

Methyl 4-(((1-Phenyl-1H-pyrazolo[3,4-d]pyrimidin-4-yl)amino)-methyl)benzoate (26b). 4-Chloro-1-phenyl-1H-pyrazolo[3,4-d]-pyrimidine (25b, 250 mg, 1.08 mmol) was diluted in THF (dry, 10 mL), and DIPEA (140 mg, 1.08 mmol) was added. In a separate flask, methyl 4-(aminomethyl)benzoate hydrochloride (17a, 219 mg, 1.08 mmol) and DIPEA (140 mg, 1.08 mmol) were dissolved in THF (dry, 10 mL) and stirred at RT for 5 min. The latter suspension was added to the first solution, and the mixture was heated to 60 °C for 20 h. EtOAc (150 mL) was added, and the organic phase was washed three times with H₂O (50 mL). The organic phase was separated and dried over MgSO₄, and the solvent was removed under reduced pressure. The crude product was purified using column chromatography on silica (cyclohexane/EtOAc 10:1) to afford 358 mg (0.996 mmol, 92%) of compound 26b as a colorless solid. TLC: R_f = 0.59 (SiO₂, cyclohexane/EtOAc 1:1). ¹H NMR (400 MHz, DMSO-d_6, 300 K): δ 9.07 (t, ³*J* = 5.6 Hz, 1H), 8.45 (s, 1H), 8.38 (s, 1H), 8.20 (d, ³*J* = 8.1 Hz, 2H), 7.95 (d, ³*J* = 8.1 Hz, 2H), 7.59−7.49 (m, 4H), 7.35 (t, ³*J* = 7.3 Hz, 1H), 4.88 (d, ³*J* = 5.8 Hz, 2H) 3.84 (s, 3H) ppm. ¹³C NMR (150 MHz, DMSO-d_6, 300 K): δ 166.0, 156.4 (2×), 152.8, 144.8, 138.9, 133.7, 129.3 (2×), 129.1, 127.5, 126.1, 120.6, 101.8, 52.2, 42.9 ppm. MS (ESI pos.) *m/z* (%): 359.89 (100) ([M − H]⁺ calcd, 360.15).

Methyl 4-((Imidazo[1,2-a]pyridine-3-carboxamido)methyl)-benzoate (26c). Imidazo[1,2-a]pyridine-3-carboxylic acid (25c, 500 mg, 3.08 mmol), HOBt (500 mg, 3.70 mmol), and EDC·HCl (709 mg, 3.70 mmol) were dissolved in DMF (dry, 8 mL), and DIPEA (598 mg, 4.63 mmol) was added. The mixture was stirred for 1 h at RT, and a solution of methyl 4-(aminomethyl)benzoate hydrochloride (17a, 789 mg, 3.91 mmol) dissolved in DMF (dry, 7 mL), DMSO (dry, 3 mL), and DIPEA (598 mg, 4.63 mmol) was added. The reaction mixture was stirred for additional 20 h at RT. DCM (200 mL) was added, and the organic phase was washed four times with H₂O (100 mL). The organic phase was separated and dried over MgSO₄, and the solvent was removed under reduced pressure. The crude product was dissolved in DCM and precipitated with cyclohexane to afford 837 mg (2.71 mmol, 88%) of compound 26c as a colorless solid. TLC: R_f = 0.26 (SiO₂, cyclohexane/EtOAc 1:3). ¹H NMR (600 MHz, DMSO-d_6, 300 K): δ 9.47 (d, ³*J* = 7.0 Hz, 1H), 9.13 (t, ³*J* = 5.8 Hz, 1H), 8.42 (s, 1H), 7.93 (d, ³*J* = 8.2 Hz, 2H), 7.72 (d, ³*J* = 9.2 Hz, 1H), 7.49 (d, ³*J* = 8.0 Hz, 2H), 7.46 (t, ³*J* = 7.7 Hz, 1H), 7.11 (t, ³*J* = 6.9 Hz, 1H), 4.60 (d, ³*J* = 5.8 Hz, 2H), 3.83 (s, 3H) ppm. ¹³C NMR (150 MHz, DMSO-d_6, 300 K): δ 166.1, 160.2, 146.9, 145.3, 136.9, 129.3, 128.2, 127.5, 127.4, 126.9, 117.9, 117.2, 113.8, 52.0, 41.6 ppm. MS (ESI neg.) *m/z* (%): 308.09 (100) ([M − H]⁻ calcd, 308.10).

Methyl 4-((5-Amino-1-methyl-1H-pyrazole-4-carboxamido)-methyl)benzoate (26d). 5-Amino-1-methyl-1H-pyrazole-4-carboxylic acid (25d, 500 mg, 3.54 mmol), HOBt (574 mg, 4.25 mmol), and EDC·HCl (815 mg, 4.25 mmol) were dissolved in DMF (dry, 10 mL), and DIPEA (687 mg, 5.31 mmol) was added. The mixture was stirred for 1 h at RT, and a solution of methyl 4-(aminomethyl)benzoate hydrochloride (17a, 857 mg, 4.25 mmol) dissolved in DMF (dry, 10 mL) and DIPEA (687 mg, 5.31 mmol) was added. The reaction mixture was stirred for additional 20 h at RT. DCM (100 mL) was added, and the organic phase was washed four times with H₂O (100 mL) and once with brine (100 mL). The organic phase was separated and dried over MgSO₄, and the solvent was removed under reduced pressure. The crude product was precipitated from cyclohexane to afford 722 mg (2.50 mmol, 71%) of compound 26d as a colorless solid. TLC: R_f = 0.25 (SiO₂, cyclohexane/EtOAc 1:3). ¹H NMR (600 MHz, DMSO-d_6, 300 K): δ 8.32 (t, ³*J* = 5.9 Hz, 1H), 7.92 (d, ³*J* = 8.1 Hz, 2H), 7.68 (s, 1H), 7.41 (d, ³*J* = 8.1 Hz, 2H), 6.16 (s, 2H), 4.45 (d, ³*J* = 6.1 Hz, 2H), 3.84 (s, 3H), 3.51 (s, 3H) ppm. ¹³C NMR (150 MHz, DMSO-d_6, 300 K): δ 166.1, 164.1, 149.4, 146.1, 136.1, 129.2, 128.0, 127.2, 96.7, 52.0, 41.3, 33.9 ppm. MS (ESI pos.) *m/z* (%): 289.03 (100) ([M + H]⁺ calcd, 289.13).

Methyl 4-(1-(5-Amino-1-methyl-1H-pyrazole-4-carboxamido)-ethyl)benzoate (26e). 5-Amino-1-methyl-1H-pyrazole-4-carboxylic acid (25d, 436 mg, 3.09 mmol), HOBt (501 mg, 3.71 mmol) and EDC·HCl (711 mg, 3.71 mmol) were dissolved in DMF (dry, 8 mL),

and DIPEA (600 mg, 4.64 mmol) was added. The mixture was stirred for 1 h at RT, and a solution of methyl 4-(1-aminoethyl)benzoate hydrochloride (17b, 800 mg, 3.71 mmol) dissolved in DMF (dry, 7 mL) and DIPEA (600 mg, 4.64 mmol) was added. The reaction mixture was stirred for additional 20 h at RT. Then, DCM (100 mL) was added, and the organic phase was washed four times with H₂O (50 mL) and once with a NaCl solution (sat., 50 mL). The organic phase was separated and dried over MgSO₄, and the solvent was removed under reduced pressure. The crude product was precipitated from cyclo-hexane to afford 632 mg (2.09 mmol, 68%) of compound 26e as a colorless solid. TLC: R_f = 0.20 (SiO₂, cyclohexane/EtOAc 1:3). ¹H NMR (600 MHz, DMSO-d_6, 300 K): δ 8.25−7.68 (m, 4H), 7.65−7.28 (m, 2H), 6.11 (s, 2H), 5.30−4.95 (m, 1H), 3.83 (s, 3H), 3.50 (s, 3H), 1.43 (s, 3H) ppm. ¹³C NMR (150 MHz, DMSO-d_6, 300 K): δ 166.1, 163.4, 151.2, 149.4, 136.3, 129.2, 127.8, 126.2, 96.7, 52.0, 47.2, 33.8, 22.1 ppm. MS (ESI neg.) *m/z* (%): 301.14 (100) ([M − H]⁻ calcd, 301.13), 347.06 (35) ([M − H + EtOH]⁻ calcd, 347.17).

Methyl 4-(1-(Imidazo[1,2-a]pyridine-3-carboxamido)ethyl)-benzoate (26f). Imidazo[1,2-a]pyridine-3-carboxylic acid (25c, 251 mg, 1.55 mmol), HOBt (209 mg, 1.85 mmol) and EDC·HCl (296 mg, 1.85 mmol) were dissolved in DMF (dry, 8 mL), and DIPEA (300 mg, 2.32 mmol) was added. The mixture was stirred for 1 h at RT, and a solution of methyl 4-(1-aminoethyl)benzoate hydrochloride (17b, 400 mg, 3.71 mmol) dissolved in DMF (dry, 7 mL), DMSO (dry, 5 mL), and DIPEA (300 mg, 2.32 mmol) was added. The reaction mixture was stirred for additional 20 h at RT; then, DCM (150 mL) was added, and the organic phase was washed four times with H₂O (100 mL). The organic phase was separated and dried over MgSO₄ and the solvent was removed under reduced pressure. The crude product was purified using column chromatography on silica (cyclohexane/EtOAc 1:10) to afford 365 mg (1.13 mmol, 73%) of compound 26f as a yellowish solid. TLC: R_f = 0.31 (SiO₂, cyclohexane/EtOAc 1:3). ¹H NMR (600 MHz, DMSO-d_6, 300 K): δ 9.40 (d, ³*J* = 7.2 Hz, 1H), 8.88 (d, ³*J* = 7.6 Hz, 1H), 8.51 (s, 1H), 7.93 (d, ³*J* = 8.3 Hz, 2H), 7.70 (d, ³*J* = 8.9 Hz, 1H) 7.55 (d, ³*J* = 8.1 Hz, 2H), 7.44 (t, ³*J* = 7.6 Hz, 1H), 7.08 (t, ³*J* = 6.9 Hz, 1H), 5.27 (p, ³*J* = 7.2 Hz, 1H), 3.83 (s, 3H), 1.52 (d, ³*J* = 7.1 Hz, 3H) ppm. ¹³C NMR (150 MHz, DMSO-d_6, 300 K): δ 166.1, 159.5, 150.4, 146.9, 137.1, 129.3, 128.1, 127.5, 126.9, 126.3, 117.9, 117.2, 113.8, 52.0, 47.7, 21.9 ppm. MS (ESI neg.) *m/z* (%): 322.08 (100) ([M − H]⁻ calcd, 322.12).

4-(((3-Phenyl-3H-[1,2,3]triazolo[4,5-d]pyrimidin-7-yl)amino)-methyl)benzoic Acid (27a). Methyl 4-(((3-phenyl-3H-[1,2,3]triazolo-[4,5-d]pyrimidin-7-yl)amino)methyl)benzoate (26a, 290 mg, 0.81 mmol) was dissolved in a mixture of THF/H₂O (20 mL, 1:1), and LiOH × H₂O (96 mg, 4.03 mmol) was added. The suspension was stirred for 4 d at RT and then acidified with HCl (aq, 10 mL, 1 M). The precipitate was filtered, washed with H₂O, and dried under reduced pressure in a vacuum oven to afford 231 mg (0.67 mmol, 83%) of compound 27a as a colorless solid. TLC: R_f = 0.60 (SiO₂, cyclohexane/EtOAc 1:10). ¹H NMR (300 MHz, DMSO-d_6, 300 K): (mixture of rotamers): δ 12.86 (s, 1H), 9.81−9.60 (m, 0.8H), 9.42−9.26 (m, 0.2H), 8.56−8.37 (m, 1H), 8.15 (d, ³*J* = 8.1 Hz, 2H), 7.91 (d, ³*J* = 7.8 Hz, 2H), 7.73−7.37 (m, 5H), 5.39−5.25 (m, 0.4H), 4.97−4.77 (m, 1.6H) ppm. ¹³C NMR (75 MHz, DMSO-d_6, 300 K): (mixture of rotamers): δ 167.1, 157.5, 154.5, 148.1, 144.0, 135.8, 129.6, 129.4, 128.5, 127.3, 127.2, 124.9, 121.4, 43.1 ppm. MS (ESI neg.) *m/z* (%): 345.07 (100) ([M − H]⁻ calcd, 345.11).

4-(((1-Phenyl-1H-pyrazolo[3,4-d]pyrimidin-4-yl)amino)methyl)-benzoic Acid (27b). Methyl 4-(((1-phenyl-1H-pyrazolo[3,4-d]-pyrimidin-4-yl)amino)methyl)benzoate (26b, 500 mg, 1.39 mmol) was dissolved in a mixture of THF/H₂O (20 mL, 1:1), and LiOH × H₂O (166 mg, 6.96 mmol) was added. The suspension was stirred for 4 d at RT and then acidified with HCl (aq, 10 mL, 1 M). The precipitate was filtered, washed with H₂O, and dried under reduced pressure in a vacuum oven to afford 455 mg (1.32 mmol, 95%) of compound 27b as a colorless solid. TLC: R_f = 0.02 (SiO₂, cyclohexane/EtOAc 1:10). ¹H NMR (300 MHz, DMSO-d_6, 300 K): δ 13.41−12.35 (br s, 1H), 9.16 (t, ³*J* = 6.1 Hz, 1H), 8.48 (s, 1H), 8.38 (s, 1H), 8.19 (d, ³*J* = 7.8 Hz, 2H), 7.92 (d, ³*J* = 8.3 Hz, 2H), 7.60−7.45 (m, 4H), 7.34 (tt, ³*J* = 7.4, 1.1 Hz, 1H), 4.86 (d, ³*J* = 5.9 Hz, 2H) ppm. ¹³C NMR (75 MHz, DMSO-d_6,

300 K): δ 167.1, 156.4 (2×), 152.8, 144.3, 138.9, 133.8, 129.5, 129.1 (2×), 127.4, 126.2, 120.7, 101.9, 43.1 ppm. MS (ESI neg.) *m/z* (%): 344.04 (100) ([M − H]⁻ calcd, 344.10).

4-((Imidazo[1,2-a]pyridine-3-carboxamido)methyl)benzoic Acid (27c). Methyl 4-((imidazo[1,2-a]pyridine-3-carboxamido)methyl)-benzoate (26c, 600 mg, 1.94 mmol) was dissolved in a mixture of THF/H₂O (20 mL, 1:1), and LiOH × H₂O (232 mg, 9.70 mmol) was added. The suspension was stirred for 3 d at RT and then acidified with HCl (aq, 10 mL, 1 M). The precipitate was filtered, washed with H₂O, and dried under reduced pressure in a vacuum oven to afford 367 mg (1.24 mmol, 64%) of compound 27c as a yellowish solid. TLC: R_f = 0.02 (SiO₂, cyclohexane/EtOAc 1:10). ¹H NMR (300 MHz, DMSO-d_6, 300 K): δ 9.89 (t, ³J = 5.9 Hz, 1H), 9.69 (d, ³J = 7.0 Hz, 1H), 9.03 (s, 1H), 8.08−7.67 (m, 4H), 7.58−7.45 (m, 3H), 4.59 (d, ³J = 5.8 Hz, 2H) ppm. ¹³C NMR (75 MHz, DMSO-d_6, 300 K): δ 167.1, 158.7, 144.1, 141.5, 133.3, 129.5, 129.4, 129.2, 127.7, 127.4, 118.8, 117.4, 113.5, 41.8 ppm. MS (ESI neg.) *m/z* (%): 294.10 (100) ([M − H]⁻ calcd, 294.09).

4-((5-Amino-1-methyl-1H-pyrazole-4-carboxamido)methyl)-benzoic Acid (27d). Methyl 4-((5-amino-1-methyl-1H-pyrazole-4-carboxamido)methyl)benzoate (26d, 600 mg, 2.08 mmol) was dissolved in a mixture of THF/H₂O (10 mL, 1:1), and LiOH × H₂O (249 mg, 10.4 mmol) was added. The suspension was stirred for 24 h at RT and then acidified with HCl (aq, 10 mL, 1 M). The precipitate was filtered, washed with H₂O, and dried under reduced pressure in a vacuum oven to afford 504 mg (1.84 mmol, 88%) of compound 27d as a colorless solid. TLC: R_f = 0.08 (SiO₂, cyclohexane/EtOAc 1:3). ¹H NMR (600 MHz, DMSO-d_6, 300 K): δ 8.50−8.44 (m, 1H), 7.90 (d, ³J = 8.2 Hz, 2H), 7.84 (s, 1H), 7.39 (d, ³J = 8.2 Hz, 2H), 6.34−4.64 (m, 3H), 4.47−4.42 (m, 2H), 3.55 (s, 3H) ppm. ¹³C NMR (150 MHz, DMSO-d_6, 300 K): δ 167.2, 163.8, 149.6, 145.5, 135.7, 129.4, 129.2, 127.1, 119.0, 97.0, 41.4, 33.8 ppm. MS (ESI neg.) *m/z* (%): 273.08 (100) ([M − H]⁻ calcd, 273.10), 309.07 (47) ([M + Cl]⁻ calcd, 308.53), 229.11 (28) ([M − COOH]⁻ calcd, 229.11).

4-(1-(5-Amino-1-methyl-1H-pyrazole-4-carboxamido)ethyl)-benzoic Acid (27e). Methyl 4-(1-(5-amino-1-methyl-1H-pyrazole-4-carboxamido)ethyl)benzoate (26e, 500 mg, 1.65 mmol) was dissolved in a mixture of THF/H₂O (10 mL, 1:1), and LiOH × H₂O (198 mg, 8.27 mmol) was added. The solvent was removed under reduced pressure. The crude product was triturated with DCM/H₂O (1:1), filtered, washed with H₂O, and dried under reduced pressure in a vacuum oven to afford 368 mg (1.28 mmol, 77%) of compound 27e as a colorless solid. TLC: R_f = 0.13 (SiO₂, DCM/MeOH 10:1). ¹H NMR (600 MHz, DMSO-d_6, 300 K): δ 12.83 (br s, 1H), 8.06 (d, ³J = 7.9 Hz, 1H), 7.89 (d, ³J = 7.9 Hz, 2H), 7.78 (s, 1H), 7.44 (d, ³J = 8.0 Hz, 2H), 6.11 (br s, 2H), 5.13 (p, ³J = 7.3 Hz, 1H), 3.50 (s, 3H), 1.43 (d, ³J = 7.2 Hz, 3H) ppm. ¹³C NMR (150 MHz, DMSO-d_6, 300 K): δ 167.2, 163.4, 150.7, 149.5, 136.3, 129.3, 129.0, 126.1, 96.7, 47.2, 33.9, 22.1 ppm. MS (ESI neg.) *m/z* (%): 287.10 (100) ([M − H]⁻ calcd, 287.29), 243.12 (20) ([M − COOH]⁻ calcd, 243.12).

4-(1-(Imidazo[1,2-a]pyridine-3-carboxamido)ethyl)benzoic Acid (27f). Methyl 4-(1-(imidazo[1,2-a]pyridine-3-carboxamido)ethyl)-benzoate (26f, 337 mg, 1.04 mmol) was dissolved in a mixture of THF/H₂O (20 mL, 1:1), and LiOH × H₂O (125 mg, 5.21 mmol) was added. The suspension was stirred for 4 d at RT and then acidified with HCl (aq, 10 mL, 1 M). The precipitate was filtered, washed with H₂O, and dried under reduced pressure in a vacuum oven to afford 308 mg (1.00 mmol, 96%) of compound 27f as a yellowish solid. TLC: R_f = 0.07 (SiO₂, cyclohexane/EtOAc 1:3). ¹H NMR (300 MHz, DMSO-d_6, 300 K): δ 12.85 (s, 1H), 9.41 (d, ³J = 7.1 Hz, 1H), 8.93 (d, ³J = 7.7 Hz, 1H), 8.55 (s, 1H), 7.91 (d, ³J = 8.3 Hz, 2H), 7.71 (d, ³J = 9.0 Hz, 1H) 7.53 (d, ³J = 8.3 Hz, 2H), 7.45 (t, ³J = 7.7 Hz, 1H), 7.09 (t, ³J = 7.0 Hz, 1H), 5.26 (p, ³J = 7.3 Hz, 1H), 1.52 (d, ³J = 7.1 Hz, 3H) ppm. ¹³C NMR (75 MHz, DMSO-d_6, 300 K): δ 167.2, 159.5, 150.4, 146.8, 137.0, 129.5, 129.3, 127.6, 127.1, 126.2, 118.0, 117.2, 113.9, 47.8, 22.0 ppm. MS (ESI neg.) *m/z* (%): 308.12 (100) ([M − H]⁻ calcd, 308.10).

N-((S)-4-Cyclohexyl-1-(((SR)-1-(methylsulfonyl)piperidin-3-yl)-amino)-1-oxobutan-2-yl)-4-(((3-phenyl-3H-[1,2,3]triazolo[4,5-d]-pyrimidin-7-yl)amino)methyl)benzamide (28). 4-(((3-Phenyl-3H-[1,2,3]triazolo[4,5-d]pyrimidin-7-yl)amino)methyl)benzoic acid (27a, 200 mg, 0.58 mmol) and HATU (263 mg, 0.69 mmol) were

dissolved in DMF (5 mL), and DIPEA (90 mg, 0.693 mmol) was added. The solution was stirred for 1.5 h at RT, and (S)-2-amino-4-cyclohexyl-N-((SR)-1-(methylsulfonyl)piperidin-3-yl)butanamide (11d, 239 mg, 0.70 mmol), diluted in DMF (5 mL), was added. The mixture was stirred for 20 h at RT. DCM (150 mL) was added, and the organic phase was washed five times with H₂O (150 mL) and dried over MgSO₄. The solvent was evaporated under reduced pressure, and the crude product was triturated with EtOAc/cyclohexane (1:1) and recrystallized from DCM to afford 51 mg (0.075 mmol, 13%) of compound 28 as a colorless solid. TLC: R_f = 0.20 (SiO₂, cyclohexane/EtOAc 1:10). HPLC: (mixture of diastereomers) $t_{R,1}$ = 15.97 min; $t_{R,2}$ = 16.27 min, purity: >95%. ¹H NMR (400 MHz, DMSO-d_6, 300 K): (mixture of diastereomers/rotamers): δ 9.77−9.63 (m, 1H) 8.56−8.39 (m, 2H), 8.22−8.09 (m, 3H), 7.72−7.60 (m, 2H), 7.58−7.39 (m, 3H), 7.32 (d, ³J = 7.5 Hz, 2H), 4.84 (d, ³J = 5.6 Hz, 2H), 3.89−3.10 (m, 4H), 2.95−2.79 (m, 5H), 1.87−1.38 (m, 12H), 1.30−1.03 (m, 5H), 0.94−0.76 (m, 2H). MS (ESI pos.) *m/z* (%): 696.30 (100) ([M + Na]⁺ calcd, 696.31). HRMS (FTMS + p MALDI) *m/z*: 696.3054 [M + Na]⁺ calcd for [C₃₄H₄₈N₉NaO₄S]⁺, 696.3056.

N-((S)-4-Cyclohexyl-1-(((SR)-1-(methylsulfonyl)piperidin-3-yl)-amino)-1-oxobutan-2-yl)-4-(((1-phenyl-1H-pyrazolo[3,4-d]-pyrimidin-4-yl)amino)methyl)benzamide (29). 4-(((1-Phenyl-1H-pyrazolo[3,4-d]pyrimidin-4-yl)amino)methyl)benzoic acid (27b, 200 mg, 0.58 mmol) and HATU (264 mg, 0.70 mmol) were dissolved in DMF (5 mL), and DIPEA (90 mg, 0.70 mmol) was added. The solution was stirred for 1.5 h at RT, and (S)-2-amino-4-cyclohexyl-N-((SR)-1-(methylsulfonyl)piperidin-3-yl)butanamide (11d, 240 mg, 0.70 mmol), diluted in DMF (5 mL), was added. The mixture was stirred for 20 h at RT. DCM (150 mL) was added, and the organic phase was washed five times with H₂O (100 mL) and dried over MgSO₄. The solvent was evaporated under reduced pressure, and the crude product was triturated with EtOAc/cyclohexane (1:1) and recrystallized from DCM to afford 125 mg (0.19 mmol, 32%) of compound 29 as a colorless solid. TLC: R_f = 0.21 (SiO₂, cyclohexane/EtOAc 1:10). HPLC: (mixture of diastereomers) $t_{R,1}$ = 9.76 min; $t_{R,2}$ = 9.92 min, purity: >95%. ¹H NMR (300 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 9.03 (t, ³J = 5.8 Hz, 1H), 8.45 (s, 1H), 8.38 (s, 1H), 8.31 (d, ³J = 8.0 Hz, 1H), 8.19 (d, ³J = 8.0 Hz, 2H), 8.00 (d, ³J = 7.6 Hz, 1H), 7.86 (d, ³J = 7.8 Hz, 2H), 7.55 (t, ³J = 7.8 Hz, 2H), 7.45 (d, ³J = 8.0 Hz, 2H), 7.34 (t, ³J = 7.6 Hz, 1H), 4.84 (d, ³J = 5.8 Hz, 2H), 4.46−4.34 (m, 1H), 3.79−3.64 (m, 1H), 3.53−3.40 (m, 1H), 3.37−3.23 (m, 1H), 2.84 (s, 3H), 2.83−2.69 (m, 1H), 2.59−2.41 (m, 1H), 1.86−1.32 (m, 11H), 1.30−1.01 (m, 6H), 0.95−0.72 (m, 2H) ppm. ¹³C NMR (150 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 171.6, 166.1, 156.4, 152.9, 142.4, 138.9, 133.7, 132.9, 129.1, 127.7, 127.0, 126.1, 120.7, 101.8, 53.4, 49.5, 49.4, 45.3, 45.2, 45.1, 43.0, 36.8, 34.4, 34.3, 33.2, 32.9, 32.6, 29.4, 28.6, 26.1, 25.7, 22.9 ppm. MS (ESI pos.) *m/z* (%): 673.25 (100) ([M + H]⁺ calcd, 673.33), 494.98 (26) ([M₆]⁺ calcd, 495.28). HRMS (FTMS + p MALDI) *m/z*: 695.3099 [M + Na]⁺ calcd for [C₃₅H₄₄N₈NaO₄S]⁺, 695.3104.

N-(4-(((S)-4-Cyclohexyl-1-(((SR)-1-(methylsulfonyl)piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamoyl)benzyl)imidazo[1,2-a]-pyridine-3-carboxamide (30). 4-((Imidazo[1,2-a]pyridine-3-carboxamido)methyl)benzoic acid (27c, 200 mg, 0.677 mmol) and HATU (309 mg, 0.813 mmol) were dissolved in DMF (5 mL), and DIPEA (105 mg, 0.813 mmol) was added. The solution was stirred for 1.5 h at RT, and (S)-2-amino-4-cyclohexyl-N-((SR)-1-(methylsulfonyl)piperidin-3-yl)butanamide (11d, 281 mg, 0.813 mmol), diluted in DMF (5 mL), was added. The mixture was stirred for 20 h at RT. DCM (150 mL) was added, and the organic phase was washed five times with H₂O (150 mL) and dried over MgSO₄. The solvent was evaporated under reduced pressure, and the crude product was purified using column chromatography on silica (EtOAc/MeOH 10:1) to afford 48 mg (0.077 mmol, 11%) of compound 30 as a colorless solid. TLC: R_f = 0.19 (SiO₂, EtOAc/MeOH 10:1). HPLC: (mixture of diastereomers) $t_{R,1}$ = 7.88 min; $t_{R,2}$ = 8.04 min, purity: 93%. ¹H NMR (300 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 9.47 (d, ³J = 7.0 Hz, 1H), 9.09 (t, ³J = 6.1 Hz, 1H), 8.40 (s, 1H), 8.31 (d, ³J = 8.0 Hz, 1H), 8.00 (d, ³J = 7.5 Hz, 1H), 7.85 (d, ³J = 8.3 Hz, 2H), 7.72 (d, ³J = 9.0 Hz, 1H), 7.51−7.31 (m, 3H), 7.12 (td, ³J = 6.9, 1.3 Hz,

1H), 4.57 (d, ^3J = 6.1 Hz, 2H), 4.45–4.33 (m, 1H), 3.79–3.65 (m, 1H), 3.53–3.37 (m, 1H), 3.36–3.25 (m, 1H), 2.85 (s, 3H), 2.83–2.70 (m, 1H), 2.59–2.41 (m, 1H), 1.88–1.31 (m, 11H), 1.29–1.00 (m, 6H), 0.93–0.71 (m, 2H) ppm. ^{13}C NMR (75 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 171.7, 166.1, 160.2, 146.9, 143.0, 136.9, 132.8, 127.6 (2×), 127.0, 126.9, 118.0, 117.3, 113.9, 107.0, 53.6, 53.4, 49.4, 45.4, 45.2, 45.1, 41.6, 36.8, 34.4, 34.3, 33.3, 32.9, 32.7, 29.4, 28.6, 26.1 (2×), 25.8, 23.0 ppm. MS (ESI neg.) m/z (%): 667.29 (100) ([M + EtOH − H]$^-$ calcd, 667.32), 621.30 (94) ([M − H]$^-$ calcd, 621.28), 361.24 (48) ([M$_f$ + EtOH + H]$^-$ calcd, 361.24). HRMS (FTMS + p MALDI) m/z: 645.2801 [M + Na]$^+$ calcd for [C$_{32}$H$_{42}$N$_7$NaO$_5$S]$^+$, 645.2835.

N-((SR)-1-(4-(((S)-4-Cyclohexyl-1-(((SR)-1-(methylsulfonyl)-piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamoyl)phenyl)ethyl)-imidazo[1,2-a]pyridine-3-carboxamide (31). 4-(1-(imidazo[1,2-a]pyridine-3-carboxamido)ethyl)benzoic acid (27f, 200 mg, 0.65 mmol) and HATU (295 mg, 0.78 mmol) were dissolved in DMF (6 mL), and DIPEA (100 mg, 0.78 mmol) was added. The solution was stirred for 1.5 h at RT, and (S)-2-amino-4-cyclohexyl-N-((SR)-1-(methylsulfonyl)piperidin-3-yl)butanamide (11d, 268 mg, 0.78 mmol), diluted in DMF (6 mL), was added. The mixture was stirred for 20 h at RT. DCM (100 mL) was added, and the organic phase was washed four times with H$_2$O (100 mL) and dried over MgSO$_4$. The solvent was evaporated under reduced pressure, and the crude product was purified using column chromatography on silica (hexane/EtOAc 1:1 → EtOAc → EtOAc/MeOH 10:1) to afford 209 mg (0.33 mmol, 51%) of compound 31 as a colorless solid. TLC: R_f = 0.20 (SiO$_2$, EtOAc/MeOH 10:1). HPLC: (mixture of diastereomers) $t_{R,1}$ = 7.96 min; $t_{R,2}$ = 8.12 min, purity: >95%. ^{1}H NMR (600 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 9.41 (d, ^3J = 6.9 Hz, 1H), 8.84 (d, ^3J = 7.9 Hz, 1H), 8.49 (d, ^4J = 1.5 Hz, 1H), 8.30 (d, ^3J = 8.1 Hz, 1H), 7.99 (d, ^3J = 7.5 Hz, 1H), 7.85 (dt, ^3J = 8.3, 3.0 Hz, 2H), 7.70 (dt, ^3J = 9.0, 1.1 Hz, 1H), 7.49 (d, ^3J = 8.3 Hz, 2H), 7.44 (ddd, ^3J = 9.0, 6.8, 1.3 Hz, 1H), 7.08 (td, ^3J = 6.9, 1.1 Hz, 1H), 5.25 (p, ^3J = 7.2 Hz, 1H), 4.43–4.36 (m, 1H), 3.76–3.69 (m, 1H), 3.51–3.41 (m, 1H), 3.36–3.29 (m, 1H), 2.84 (s, 3H), 2.79–2.71 (m, 1H), 2.57–2.52 (m, 1H), 1.84–1.55 (m, 10H), 1.52 (d, ^3J = 7.1 Hz, 3H), 1.44–1.31 (m, 1H), 1.28–1.06 (m, 6H), 0.89–0.77 (m, 2H) ppm. ^{13}C NMR (150 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 171.6, 166.1, 159.4, 148.2, 146.9, 137.0, 132.7, 127.6, 127.5, 126.9, 125.8, 118.0, 117.2, 113.8, 53.5, 53.4, 49.6, 49.4, 47.7, 45.4, 45.2, 45.1, 36.8 (2×), 34.4, 34.3, 33.2, 32.9, 32.6, 29.4, 28.9, 28.6, 26.1, 25.8, 25.7, 23.0 (2×), 22.2, 22.1 ppm. MS (ESI pos.) m/z (%): 637.50 (100) ([M + H]$^+$ calcd, 637.32), 459.31 (60) ([M$_f$]$^+$ calcd, 459.24). HRMS (FTMS + p MALDI) m/z: 637.3173 [M + H]$^+$ calcd for [C$_{33}$H$_{45}$N$_6$O$_5$S]$^+$, 637.3172.

5-Amino-N-(4-(((S)-4-cyclohexyl-1-(((SR)-1-(methylsulfonyl)-piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamoyl)benzyl)-1-methyl-1H-pyrazole-4-carboxamide (32). 4-((5-Amino-1-methyl-1H-pyrazole-4-carboxamido)methyl)benzoic acid (27d, 200 mg, 0.73 mmol) and HATU (333 mg, 0.88 mmol) were dissolved in DMF (5 mL), and DIPEA (113 mg, 0.88 mmol) was added. The solution was stirred for 1.5 h at RT; (S)-2-amino-4-cyclohexyl-N-((SR)-1-(methylsulfonyl)-piperidin-3-yl)butanamide (11d, 302 mg, 0.875 mmol), diluted in DMF (5 mL), was added. The mixture was stirred for 20 h at RT. DCM (100 mL) was added, and the organic phase was washed four times with H$_2$O (100 mL) and dried over MgSO$_4$. The solvent was evaporated under reduced pressure; the crude product was triturated with EtOAc/cyclohexane (1:1) and recrystallized from DCM to afford 281 mg (0.47 mmol, 64%) of compound 32 as a colorless solid. TLC: R_f = 0.37 (SiO$_2$, DCM/MeOH 10:1). HPLC: (mixture of diastereomers) $t_{R,1}$ = 8.17 min; $t_{R,2}$ = 8.34 min, purity: 94%. ^{1}H NMR (300 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 8.35–8.24 (m, 2H), 8.01 (d, ^3J = 7.6 Hz, 1H), 7.83 (d, ^3J = 7.2 Hz, 2H), 7.68 (s, 1H), 7.36–7.28 (m, 2H), 6.15 (s, 2H), 4.46–4.35 (m, 3H), 3.77–3.67 (m, 1H), 3.51 (s, 3H), 3.50–3.40 (m, 1H), 3.36–3.28 (m, 1H), 2.90–2.71 (m, 4H), 2.58–2.47 (m), 1.82–1.48 (m, 10H), 1.45–1.31 (m, 1H), 1.28–1.03 (m, 6H), 0.90–0.79 (m, 2H) ppm. ^{13}C NMR (75 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 171.6 (2×), 166.1 (2×), 164.1, 149.4, 143.8, 136.1, 132.6, 132.5, 127.5, 127.0, 126.7, 126.6, 96.7, 53.5, 53.4, 49.6, 49.4, 45.4, 45.2, 45.1, 41.2, 36.8, 34.4, 34.3, 33.9, 33.2, 32.9, 32.8,

32.7, 32.6, 29.4, 28.9, 28.6, 26.1, 25.8 (2×), 23.0, 22.9 ppm. MS (ESI neg.) m/z (%): 646.16 (39) ([M + EtOH − H]$^-$ calcd, 646.33), 600.19 (100) ([M − H]$^-$ calcd, 600−29). HRMS (FTMS + p MALDI) m/z: 624.2950 [M + Na]$^+$ calcd for [C$_{29}$H$_{43}$N$_7$NaO$_5$S]$^+$, 624.2944.

5-Amino-N-((SR)-1-(4-(((S)-4-cyclohexyl-1-(((SR)-1-(methylsulfonyl)piperidin-3-yl)amino)-1-oxobutan-2-yl)carbamoyl)phenyl)ethyl)-1-methyl-1H-pyrazole-4-carboxamide (33). 4-(1-(5-Amino-1-methyl-1H-pyrazole-4-carboxamido)ethyl)-benzoic acid (27e, 125 mg, 0.40 mmol) and HATU (184 mg, 0.49 mmol) were dissolved in DMF (5 mL), and DIPEA (63 mg, 0.40 mmol) was added. The solution was stirred for 1.5 h at RT; (S)-2-amino-4-cyclohexyl-N-((SR)-1-(methylsulfonyl)piperidin-3-yl)-butanamide (11d, 168 mg, 0.49 mmol), diluted in DMF (5 mL), was added. The mixture was stirred for 20 h at RT. DCM (100 mL) was added, and the organic phase was washed three times with H$_2$O (100 mL) and once with a NaCl solution (sat., 50 mL) and dried over MgSO$_4$. The solvent was evaporated under reduced pressure, and the crude product was purified using flash column chromatography on silica (cyclohexane/EtOAc 1:1 → EtOAc → EtOAc/MeOH 10:1) to afford 162 mg (0.26 mmol, 65%) of compound 33 as a colorless solid. TLC: R_f = 0.24 (SiO$_2$, EtOAc/MeOH 10:1). HPLC: (mixture of diastereomers) $t_{R,1}$ = 8.34 min; $t_{R,2}$ = 8.51 min, purity: >95%. ^{1}H NMR (600 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 8.28 (d, ^3J = 8.1 Hz, 1H), 8.02 (d, ^3J = 7.8 Hz, 1H), 7.99 (d, ^3J = 8.0 Hz, 1H), 7.82 (dt, ^3J = 8.4, 2.6 Hz, 2H), 7.77 (s, 1H), 7.40 (d, ^3J = 8.4 Hz, 2H), 6.10 (s, 2H), 5.11 (p, ^3J = 7.3 Hz, 1H), 4.43–4.35 (m, 1H), 3.80–3.68 (m, 1H), 3.52–3.47 (m, 1H), 3.49 (s, 3H), 3.46–3.42 (m, 1H), 2.86–2.68 (m, 4H), 2.57–2.48 (m, 1H), 1.84–1.48 (m, 10H), 1.42 (d, ^3J = 7.1 Hz, 3H), 1.45–1.30 (m, 1H), 1.28–1.05 (m, 6H), 0.92–0.78 (m, 2H) ppm. ^{13}C NMR (150 MHz, DMSO-d_6, 300 K): (mixture of diastereomers): δ 171.6 (2×), 166.2, 166.1, 163.4, 149.4, 148.9, 136.2, 132.4, 127.4, 125.7, 113.7, 96.7, 53.5, 53.4, 49.5, 49.4, 49.2, 47.1, 45.4 (2×), 45.2, 45.1, 36.9, 36.8, 34.4, 34.3, 33.8, 33.2, 32.9, 32.6, 29.4, 28.9, 28.6, 26.1 (2×), 26.1 (2×), 25.8 (2×), 25.7, 23.0 (2×), 22.3, 22.2 ppm. MS (ESI pos.) m/z (%): 616.22 (100) ([M + H]$^+$ calcd, 616.33), 438.14 (37) ([M$_f$]$^+$ calcd, 438.25). HRMS (FTMS + p MALDI) m/z: 638.3092 [M + Na]$^+$ calcd for [C$_{30}$H$_{45}$N$_7$NaO$_5$S]$^+$, 638.3101.

***rac-tert*-Butyl (1-(Methylsulfonyl)piperidin-3-yl)carbamate (36a).** *rac-tert*-Butyl piperidin-3-ylcarbamate (34a, 801 mg, 4.00 mmol) was dissolved in DCM (dry, 20 mL), and triethylamine (405 mg, 4.00 mmol) was added. The solution was cooled to −15 °C, and methanesulfonyl chloride (35a, 687 mg, 6.00 mmol) was added dropwise. The mixture was stirred for 6 h at RT. The reaction mixture was acidified with NH$_4$Cl (sat, 3 mL) and extracted three times with DCM. The separated organic phase was washed three times with H$_2$O (20 mL) and dried over MgSO$_4$. The solvent was evaporated under reduced pressure to afford 1.08 mg (3.88 mmol, 97%) of 36a as a colorless solid. TLC: R_f = 0.63 (SiO$_2$, cyclohexane/EtOAc, 1:20, ninhydrin). ^{1}H NMR (500 MHz, CDCl$_3$, 300 K): δ 4.95–4.78 (m, 1H), 3.89–3.74 (m, 1H), 3.72–2.95 (m, 4H), 2.77 (s, 3H), 1.88–1.57 (m, 4H), 1.43 (s, 9H). ^{13}C NMR (126 MHz, CDCl$_3$, 300 K): δ 155.0, 79.8, 50.8, 46.3, 45.6, 35.2, 29.2, 28.5, 22.1 ppm. MS (ESI neg.) m/z (%): 323.08 (84) ([M + EtOH − H]$^-$ calcd, 323.16), 313.06 (100) ([M + Cl]$^-$ calcd, 313.58), 162.95 (67) ([M$_f$]$^-$ calcd, 162.95), 160.92 (60) ([M$_f$ − H]$^-$ calcd, 160.94).

***tert*-Butyl (S)-(1-(Methylsulfonyl)piperidin-3-yl)carbamate (36b).** The synthesis of *tert*-butyl (S)-(1-(methylsulfonyl) piperidin-3-yl)carbamate (36b) was performed in analogy to that of compound 36a. *tert*-Butyl (S)-piperidin-3-ylcarbamate (34b, 801 mg, 4.00 mmol) was used as a starting material to afford 1.03 g (3.71 mmol, 93%) of compound 36b as a colorless solid. TLC: R_f = 0.63 (SiO$_2$, cyclohexane/EtOAc, 1:20, ninhydrin). ^{1}H NMR (250 MHz, CDCl$_3$, 300 K): δ 6.95 (d, ^3J = 7.5 Hz, 1H), 3.58–3.46 (m, 1H) 3.40–3.01 (m, 2H), 2.85 (s, 3H), 2.71–2.57 (m, 1H), 2.47–2.35 (m, 1H), 1.84–1.67 (m, 2H), (s, 9H), 1.59–1.11 (m, 2H) ppm. ^{13}C NMR (126 MHz, CDCl$_3$, 300 K): δ 155.0, 79.8, 50.8, 46.3, 45.6, 35.2, 29.2, 28.5, 22.1 ppm. MS (ESI neg.) m/z (%): 323.15 (100) ([M + EtOH − H]$^-$ calcd, 323.16), 313.13 (39) ([M + Cl]$^-$ calcd, 313.58).

***tert*-Butyl (R)-(1-(Methylsulfonyl)piperidin-3-yl)carbamate (36c).** The synthesis of *tert*-butyl (R)-(1-(methylsulfonyl)piperidin-3-yl)-

carbamate (36c) was performed analogously to that of compound 36b. *tert*-Butyl (R)-piperidin-3-ylcarbamate (34c, 801 mg, 4.00 mmol) was used as a starting material to afford 915 mg (3.29 mmol, 97%) of compound 36c as a light-brownish solid. The analytical data are in accordance with those of compound 36b.

rac-tert-Butyl (1-(Cyclopropanecarbonyl)piperidin-3-yl)-carbamate (36d). Cyclopropanecarboxylic acid (35b, 215 mg, 2.50 mmol) and HATU (1.14 g, 3.00 mmol) were dissolved in DMF (dry, 7 mL), and DIPEA (387 mg, 3.00 mmol) was added. The reaction mixture was stirred for 1.5 h at RT, and *rac-tert*-butyl piperidin-3-ylcarbamate (34a, 600 mg, 3.00 mmol) dissolved in DMF (dry, 5 mL) was added. The mixture was stirred for additional 20 h at RT. DCM (100 mL) was added, and the organic phase was washed four times with H_2O (100 mL) and once with NH_4Cl (sat. 100 mL) and dried over $MgSO_4$. The solvent was evaporated under reduced pressure, and the crude product was purified using column chromatography on silica (cyclohexane/EtOAc 1:9). Compound 36d was obtained in a yield of 636 mg (2.37 mmol, 95%) as a colorless solid. TLC: R_f = 0.56 (SiO_2, cyclohexane/EtOAc, 1:9, ninhydrin). ^{1}H NMR (500 MHz, CDCl₃, 300 K): δ 4.73−4.54 (m, 1H), 4.05−2.84 (m, 5H), 1.97−1.86 (m, 1H), 1.81−1.49 (m, 4H), 1.42 (s, 9H), 1.01−0.89 (m, 2H), 0.80−0.66 (m, 2H) ppm. ^{13}C NMR (126 MHz, CDCl₃, 300 K): δ 172.7, 155.3, 79.7, 50.4, 46.9, 42.7, 30.3, 28.8, 22.5, 11.0, 7.7 ppm. MS (ESI pos.) *m/z* (%): 307.17 (23) ([M + K]⁺ calcd, 307.27), 291.20 (100) ([M + Na]⁺ calcd, 291.17), 269.21 (5) ([M + H]⁺ calcd, 269.19), 169.12 (32) ([M − Boc + H]⁺ calcd, 169.14).

rac-tert-Butyl (1-(Oxetan-3-yl)piperidin-3-yl)carbamate (36e). *rac-tert*-Butyl piperidin-3-ylcarbamate (34a, 400 mg, 2.00 mmol) was dissolved in THF (dry, 10 mL), and oxetan-3-one (35c, 107 μL, 1.66 mmol) was added at 0 °C. The mixture was stirred for 1.5 h and allowed to warm to RT. Sodium triacetoxyborohydride (708 mg, 3.33 mmol) was added, and the suspension was stirred for additional 20 h at RT. The reaction was quenched with NaHCO₃ (sat., 30 mL), and the solvent was evaporated under reduced pressure. The residue was dissolved in EtOAc (30 mL); H_2O (30 mL) was added and extracted four times with EtOAc (30 mL). The organic phase was dried over $MgSO_4$. The solvent was evaporated under reduced pressure, and the crude product was purified using column chromatography on silica (cyclohexane/EtOAc 1:20 + 1% TEA). Compound 36e was obtained in a yield of 240 mg (0.937 mmol, 47%) as a colorless solid. TLC: R_f = 0.20 (SiO_2, cyclohexane/EtOAc + 1% TEA, 1:20, ninhydrin). ^{1}H NMR (500 MHz, CDCl₃, 300 K): δ 5.16−4.92 (m, 1H, NH), 4.66−4.50 (m, 4H), 3.82−3.69 (m, 1H), 3.41 (p, ^3J = 6.4 Hz, 1H), 2.47−1.83 (m, 4H), 1.81−1.33 (m, 4H), 1.43 (s, 9H) ppm. ^{13}C NMR (126 MHz, CDCl₃, 300 K): δ 155.2, 79.3, 75.9, 59.1, 55.2, 50.2, 45.8, 29.6, 28.6, 21.9 ppm. MS (ESI neg.) *m/z* (%): 255.20 (100) ([M − H]⁻ calcd, 255.17).

rac-1-(Methylsulfonyl)piperidin-3-amine (37a). *rac-tert*-Butyl (1-(methylsulfonyl)piperidin-3-yl)carbamate (36a, 450 mg, 1.62 mmol) was dissolved in DCM (10 mL) and cooled to 0 °C, and TFA (2 mL) was added dropwise. The mixture was stirred for 1 h, allowing to warm up to RT. The solvent was removed under reduced pressure, and the residue was dissolved in MeOH. The mixture was cooled to 5 °C; K_2CO_3 (3.6 g, 26.0 mmol) was added and stirred for 10 min. The solid was filtered off, and the solvent was removed under reduced pressure. The quantitatively obtained crude product was used without further purification in the next step.

(S)-1-(Methylsulfonyl)piperidin-3-amine (37b). The synthesis of (S)-1-(methylsulfonyl)piperidin-3-amine (37b) was performed analogously to that of compound 37a. *tert*-Butyl (S)-(1-(methylsulfonyl)-piperidin-3-yl)carbamate (36b, 420 mg, 1.51 mmol) was used as a starting material to afford compound 37b in quantitative yield. The product was used without further purification in the next step.

(R)-1-(Methylsulfonyl)piperidin-3-amine (37c). The synthesis of (R)-1-(methylsulfonyl)piperidin-3-amine (37c) was performed analogously to that of compound 37b. The starting material was *tert*-butyl (R)-(1-(methylsulfonyl)piperidin-3-yl)carbamate (36c, 420 mg, 1.51 mmol). The crude product was obtained in quantitative yield and used without further purification in the next step.

rac-(3-Aminopiperidin-1-yl) (cyclopropyl)methanone (37d). The synthesis of *rac-*(3-aminopiperidin-1-yl) (cyclopropyl)methanone

(37d) was performed analogously to that of compound 37a. *rac-tert*-Butyl (1-(cyclopropanecarbonyl)piperidin-3-yl)carbamate (36d, 430 mg, 1.60 mmol) was used as a starting material to afford compound 37d in quantitative yield. The crude product was used without further purification in the next step.

rac-1-(Oxetan-3-yl)piperidin-3-amine (37e). The synthesis of *rac-*1-(oxetan-3-yl)piperidin-3-yl)piperidin-3-amine (37e) was performed analogously to that of compound 37d. *rac-tert*-Butyl (1-(oxetan-3-yl)piperidin-3-yl)carbamate (36e, 350 mg, 1.37 mmol) was used as a starting material to afford compound 37e in quantitative yield. The crude product was used without further purification in the next step.

Biological and Biochemical Methods. *DSF Assay.* Differences in the melting temperature (ΔT_m) data were measured as described in Fedorov et al.[30]

NanoBRET Assay. The assay was performed as described previously.[31,32] In brief, full-length kinase ORF (Promega) cloned in frame with a NanoLuc-vector (as indicated in Table S7) was transfected into HEK293T cells using FuGENE HD (Promega, E2312), and proteins were allowed to express for 20 h. A serially diluted inhibitor and NanoBRET kinase tracer at the respective previously determined tracer $K_{d,app}$ (as indicated in Table S7) were pipetted into white 384-well plates (Greiner 781 207) using an ECHO 550 acoustic dispenser (Labcyte). The corresponding transfected cells were added and reseeded at a density of 2×10^5 cells/mL after trypsinization and resuspension in Opti-MEM without phenol red (Life Technologies). The system was allowed to equilibrate for 2 h at 37 °C and 5% CO_2 prior to BRET measurements. To measure BRET, NanoBRET NanoGlo substrate + extracellular NanoLuc inhibitor (Promega, N2160) was added as per the manufacturer's protocol, and filtered luminescence was measured on a PHERAstar plate reader (BMG Labtech) equipped with a luminescence filter pair [450 nm BP filter (donor) and 610 nm LP filter (acceptor)]. Competitive displacement data were then plotted using GraphPad Prism 8 software using a 4-parameter curve fit with the following equation: Y = bottom + (top-bottom)/(1 + 10((log IC₅₀ − X)*HillSlope))

Protein Purification and Crystallography. Murine MAP kinase p38α was expressed in *Escherichia coli* and purified as previously described.[33] The kinase domain of human DDR1 (residues 601−913) was expressed in baculovirus and purified following the published procedure.[9] Crystals of p38α and DDR1 inhibitor complexes were grown in 3-well crystallization plates (SwisSCI) using the sitting drop vapor diffusion method. Recombinant p38α (10−13 mg/mL) was incubated with 1 mM corresponding inhibitor, and the complexes were crystallized at 4 °C using PEG smear-based conditions containing either 17.0−22.5% medium-molecular weight PEG smears and 0.1 M MES, pH 6.0−6.5, or 14% high-molecular weight PEG smears and 0.1 M MES, pH 6.0. For the DDR1-13 complex, the protein (10 mg/mL DDR1 in 50 mM HEPES, pH 7.5, 300 mM NaCl, and 1 mM TCEP) was mixed with 13 at a final inhibitor concentration of 1 mM. Crystals were then grown at 20 °C with a reservoir solution consisting of 20% PEG 3350, 10% ethylene glycol, 0.1 M bis-tris propane, pH 6.5, and 0.2 M sodium sulfate (50 nL of inhibitor and 100 nL of reservoir). Crystals of the DDR1-30 complex were grown at 4 °C by mixing the protein solution (100 nL, 16 mg/mL DDR1, 1 mM 30) with 100 nL of precipitant solution consisting of 20% PEG 10000 and 0.1 M HEPES, pH 7.5. Crystals of the inhibitor complexes were cryoprotected with mother liquor complemented with 20−25% ethylene glycol and flash-frozen in liquid nitrogen. X-ray diffraction data were collected at 100 K at Diamond beamlines i04 and i04-1, BESSY beamline 14.1, and Swiss Light Source (SLS) beamline X06SA. The data were processed with either Xia2[34] or XDS[35] and Aimless.[36] The structures were then solved by molecular replacement using Phaser[37] and published p38α (PDB ID: 5LAR) or DDR1 (PDB IDs: 4BKJ and 5FDX) coordinates as a search model. The structural models were then manually rebuilt in COOT[38] and refined using either REFMAC[39] or PHENIX.[40] The data collection and refinement statistics are summarized in Table S8. Structure figures were generated using PyMOL (https://pymol.org/2/).

Live Cell High-Content Screen. To assess the influence of the compounds on cell viability and the potential effect on microtubules, a

live cell high-content screen was performed, using a confocal microscope (CQ1, Yokogawa). HEK293T (ATCC CRL-1573) and U2OS (ATCCHTB-96) were cultured in DMEM plus L-glutamine (high glucose) supplemented with 10% FBS (Gibco) and penicillin/ streptomycin (Gibco). MRC-9 fibroblasts (ATCC CCL-2) were cultured in EMEM plus L-glutamine supplemented with 10% FBS (Gibco) and penicillin/streptomycin (Gibco). For every cell line, 2000 cells per well were seeded in 384-well plates in culture medium (cell culture microplate, PS, f-bottom, μClear, 781091, and Greiner). Cells were stained with 75 nM Hoechst 33342 (Thermo Scientific), 75 nM MitoTracker red (Invitrogen), 0.3 μl/well Annexin V Alexa Fluor 680 conjugate (Invitrogen), and 25 nL/well BioTracker 488 green microtubule cytoskeleton dye (EMD Millipore). Each compound (28, 30, 38, and 39) was tested at two different concentrations (1 and 10 μM) in triplicate. Staurosporine, 10 μM, was used as a positive control. Cellular shape and fluorescence were measured before compound treatment and 12, 24, and 48 h after compound exposure using the CQ1 high-content confocal microscope (Yokogawa). The following setup parameters were used for image acquisition: Ex 405 nm/Em 447/60 nm, 500 ms, 50%; Ex 561 nm/Em 617/73 nm, 100 ms, 40%; Ex 488/Em 525/50 nm, 50 ms, 40%; and bright field, 300 ms, 100% transmission, one centered field per well, 7 z stacks per well with 55 μm spacing. Images were analyzed using CellPathfinder software (Yokogawa). Cells were detected as described previously,[41] gated using a machine learning algorithm, and divided into the categories healthy, apoptotic, lysed, and dead cells. The program was trained using a training set of five compounds with known effects (staurosporine, paclitaxel, digitonin, milciclib, and daunorubicin). Data were normalized against the average of DMSO (0.1%)-treated cells. Significance was shown using a two-way ANOVA analysis. Two biological duplicates were performed.

TNF-α Release Assay in Human Whole Blood. The assay was performed as published by Bauer *et al.*[42] The assay principle is summarized here in brief: the tested compounds were preincubated in human whole blood from two different donors for 15 min before TNF-α release was stimulated by addition of LPS. After stimulation, the samples were incubated at 37 °C and 6% CO_2 for 2.5 h followed by centrifugation to separate cells and plasma. The supernatant was separated, and the TNF-α levels were determined using a sandwich ELISA.

In Vitro Metabolism Studies. Pooled male HLMs were purchased from Xenotech. These microsomes were characterized in protein and cytochrome P-450 content. All incubations were made in the presence of an NADPH-regenerating system, consisting of 5 mM glucose-6-phosphate, 5 U/mL glucose-6-phosphate dehydrogenase, and 1 mM NADP+. The substrate (100 μM), the NADPH regenerating system, and 3.8 mM $MgCl_2 \times 6 H_2O$ in 0.1 M Tris buffer (pH 7.4 at 37 °C) were pre-incubated for 5 min in a shaking heating block at 37 °C and 650 rpm. The incubation mix was split into 50 μL aliquots, and the reaction was started by addition of the HLMs. The microsomal protein content was standardized to 1 mg/mL. To follow the course of metabolism, the reaction tubes were quenched at selected time points (0, 10, 20, 30, 60, 120, 180, and 240 min) by adding 100 μL of ice-cold internal standard at a concentration of 20 μM in ACN. The samples were vortexed for 30 s and centrifuged (19,800 relative centrifugal force/4 °C/20 min). The supernatant was used directly for LC–MS analysis. All incubations were conducted in triplicate, and incubations with heat-inactivated HLMs were used to show that analyte reduction results from metabolic degradation only. In all incubations, a limit of 1% organic solvent was not exceeded.

LC–MS Analysis. Chromatographic separation was performed on an Alliance 2695 Separations Module (Waters GmbH, Eschborn) using a Dr. Maisch Nucleosil 100 C18, 5 μm, 53 × 4.6 mm column with water (A) and ACN (B) with 0.1% (v/v) formic acid as mobile phases. The following gradient was applied: 20% B from 0 to 0.2 min, 20 to 80% B from 0.2 to 5.5 min, and 20% B from 5.6 to 14 min at a flow rate of 400 μL/min. Samples were maintained at 4 °C, the column temperature was set to 40 °C, and the injection volume was 10 μL. The detection was performed on a Waters Quattro Micro API mass spectrometer (Waters GmbH, Eschborn) operated in the electrospray ionization (+) mode.

The spray voltage was set to 4 kV. The heated capillary operated at 270 °C, and the desolvation gas flow was set to 650 L/h.

■ ASSOCIATED CONTENT

⬤ Supporting Information
The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jmedchem.1c00868.

> Biochemical assay, selectivity data, X-ray data collection and refinement statistics as well as NMR and HRMS spectra (PDF)

> Molecular formula strings (CSV)

Accession Codes
Coordinates and structure factors of the p38α and DDR1 inhibitor complexes were deposited in the Protein Data Bank (PDB). Accession codes: 7BE4 (p38α-13), 7BE5 (p38α-32), 7BDQ (p38α-29), 7BDO (p38α-30), 7BE6 (DDR1-13), and 7BCM (DDR1-30). Authors will release the atomic coordinates upon article publication.

■ AUTHOR INFORMATION

Corresponding Author
Stefan Knapp − *Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Structural Genomics Consortium (SGC), Buchmann Institute for Molecular Life Sciences, 60438 Frankfurt am Main, Germany;* ⬤ orcid.org/0000-0001-5995-6494; Email: knapp@pharmchem.uni-frankfurt.de

Authors
Sandra Röhm − *Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Structural Genomics Consortium (SGC), Buchmann Institute for Molecular Life Sciences, 60438 Frankfurt am Main, Germany;* ⬤ orcid.org/0000-0003-3999-712X

Benedict-Tilman Berger − *Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Structural Genomics Consortium (SGC), Buchmann Institute for Molecular Life Sciences, 60438 Frankfurt am Main, Germany;* ⬤ orcid.org/0000-0002-3314-2617

Martin Schröder − *Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Structural Genomics Consortium (SGC), Buchmann Institute for Molecular Life Sciences, 60438 Frankfurt am Main, Germany;* ⬤ orcid.org/0000-0001-7631-7766

Deep Chatterjee − *Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Structural Genomics Consortium (SGC), Buchmann Institute for Molecular Life Sciences, 60438 Frankfurt am Main, Germany*

Sebastian Mathea − *Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Structural Genomics Consortium (SGC), Buchmann Institute for Molecular Life Sciences, 60438 Frankfurt am Main, Germany;* ⬤ orcid.org/0000-0001-8500-4569

Andreas C. Joerger − *Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Structural Genomics Consortium (SGC), Buchmann Institute for Molecular Life Sciences, 60438*

Frankfurt am Main, Germany; orcid.org/0000-0002-1232-0138

Daniel M. Pinkas − *Centre for Medicines Discovery, University of Oxford, Oxford OX3 7DQ, U.K.;* orcid.org/0000-0002-4740-6695

Joshua C. Bufton − *Centre for Medicines Discovery, University of Oxford, Oxford OX3 7DQ, U.K.*

Amelie Tjaden − *Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Structural Genomics Consortium (SGC), Buchmann Institute for Molecular Life Sciences, 60438 Frankfurt am Main, Germany*

Lohitesh Kovooru − *Structural Genomics Consortium (SGC), Buchmann Institute for Molecular Life Sciences, 60438 Frankfurt am Main, Germany; Institute of Biochemistry II, Faculty of Medicine, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany*

Mark Kudolo − *Department of Pharmaceutical/Medicinal Chemistry, Eberhard Karls University Tübingen, 72076 Tübingen, Germany*

Christian Pohl − *Structural Genomics Consortium (SGC), Buchmann Institute for Molecular Life Sciences, 60438 Frankfurt am Main, Germany; Institute of Biochemistry II, Faculty of Medicine, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Present Address: Neuroscience Discovery, AbbVie Deutschland GmbH & Co KG, Knollstraße 50, 67061 Ludwigshafen, Germany*

Alex N. Bullock − *Centre for Medicines Discovery, University of Oxford, Oxford OX3 7DQ, U.K.;* orcid.org/0000-0001-6757-0436

Susanne Müller − *Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany; Structural Genomics Consortium (SGC), Buchmann Institute for Molecular Life Sciences, 60438 Frankfurt am Main, Germany;* orcid.org/0000-0003-2402-4157

Stefan Laufer − *Department of Pharmaceutical/Medicinal Chemistry, Eberhard Karls University Tübingen, 72076 Tübingen, Germany;* orcid.org/0000-0001-6952-1486

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jmedchem.1c00868

Author Contributions

S.R. wrote the manuscript, did synthesis, and contributed to experimental data; structural studies were performed by M.S., A.C.J., D.C., S.M., D.M.P., and J.C.B.; NanoBRET was performed by B.-T.B., and further cellular assays were performed by A.T., M.K., and L.K.; C.P., A.N.B., S.M., S.L., and S.K. supervised research. All authors edited and approved the manuscript.

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

Abl1, abelson tyrosine protein kinase 1; CSF1R, macrophage colony-stimulating factor 1 receptor; DDR, discoidin domain receptor; DIPEA, N,N-diisopropylethylamine; DSF, differential scanning fluorimetry; ΔT_m, differences in the melting temperature; EDC−HCl, 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide−hydrochloride; EGFR, epidermal growth factor receptor; HATU, N-[(dimethylamino)-1H-1,2,3-triazolo-[4,5-b]pyridin-1-ylmethylene]-N-methylmethanaminium hexafluorophosphate N-oxide; HLM, human liver microsomes; HOBt, hydroxybenzotriazole; IRK, insulin receptor kinase; KIT, mast/stem cell growth factor receptor kit; MAPK, mitogen-activated protein kinase; MEK5, dual-specificity mitogen-activated protein kinase kinase 5; PDGFR, platelet-derived growth factor receptor; Pyk2, protein-tyrosine kinase 2-β; RTK, receptor tyrosine kinase; TEA, triethylamine; Tie, tyrosine protein kinase receptor tie; TRK, tropomyosin receptor kinase; YSK4, mitogen-activated protein kinase kinase kinase 19; ZAK, mitogen-activated protein kinase kinase kinase 20

■ REFERENCES

(1) Shrivastava, A.; Radziejewski, C.; Campbell, E.; Kovac, L.; McGlynn, M.; Ryan, T. E.; Davis, S.; Goldfarb, M. P.; Glass, D. J.; Lemke, G.; Yancopoulos, G. D. An orphan receptor tyrosine kinase family whose members serve as nonintegrin collagen receptors. *Mol. Cell* 1997, 1, 25−34.

(2) Vogel, W.; Gish, G. D.; Alves, F.; Pawson, T. The discoidin domain receptor tyrosine kinases are activated by collagen. *Mol. Cell* 1997, 1, 13−23.

(3) Li, Y.; Lu, X.; Ren, X.; Ding, K. Small molecule discoidin domain receptor kinase inhibitors and potential medical applications. *J. Med. Chem.* 2015, 58, 3287−3301.

(4) Leitinger, B. Molecular analysis of collagen binding by the human discoidin domain receptors, ddr1 and ddr2: Identification of collagen binding sites in ddr2. *J. Biol. Chem.* 2003, 278, 16761−16769.

(5) Agarwal, G.; Smith, A. W.; Jones, B. Discoidin domain receptors: Micro insights into macro assemblies. *Biochim. Biophys. Acta, Mol. Cell Res.* 2019, 1866, 118496.

(6) Leitinger, B. Chapter two - discoidin domain receptor functions in physiological and pathological conditions. In *International Review of Cell and Molecular Biology*; Jeon, K. W., Ed.; Academic Press, 2014; Vol. 310, pp 39−87.

(7) Yeung, D. A.; Shanker, N.; Sohail, A.; Weiss, B. A.; Wang, C.; Wellmerling, J.; Das, S.; Ganju, R. K.; Miller, J. L. C.; Herr, A. B.; Fridman, R.; Agarwal, G. Clustering, spatial distribution, and phosphorylation of discoidin domain receptors 1 and 2 in response to soluble collagen i. *J. Mol. Biol.* 2019, 431, 368−390.

(8) Artim, S. C.; Mendrola, J. M.; Lemmon, M. A. Assessing the range of kinase autoinhibition mechanisms in the insulin receptor family. *Biochem. J.* 2012, 448, 213−220.

(9) Canning, P.; Tan, L.; Chu, K.; Lee, S. W.; Gray, N. S.; Bullock, A. N. Structural mechanisms determining inhibition of the collagen

receptor ddr1 by selective and multi-targeted type ii kinase inhibitors. *J. Mol. Biol.* 2014, *426*, 2457–2470.

(10) Hanson, S. M.; Georghiou, G.; Thakur, M. K.; Miller, W. T.; Rest, J. S.; Chodera, J. D.; Seeliger, M. A. What makes a kinase promiscuous for inhibitors? *Cell Chem. Biol.* 2019, *26*, 390–399.

(11) Richters, A.; Nguyen, H. D.; Phan, T.; Simard, J. R.; Grütter, C.; Engel, J.; Rauh, D. Identification of type ii and iii ddr2 inhibitors. *J. Med. Chem.* 2014, *57*, 4252–4262.

(12) Murray, C. W.; Berdini, V.; Buck, I. M.; Carr, M. E.; Cleasby, A.; Coyle, J. E.; Curry, J. E.; Day, J. E. H.; Day, P. J.; Hearn, K.; Iqbal, A.; Lee, L. Y. W.; Martins, V.; Mortensen, P. N.; Munck, J. M.; Page, L. W.; Patel, S.; Roomans, S.; Smith, K.; Tamanini, E.; Saxty, G. Fragment-based discovery of potent and selective ddr1/2 inhibitors. *ACS Med. Chem. Lett.* 2015, *6*, 798–803.

(13) Liu, L.; Hussain, M.; Luo, J.; Duan, A.; Chen, C.; Tu, Z.; Zhang, J. Synthesis and biological evaluation of novel dasatinib analogues as potent ddr1 and ddr2 kinase inhibitors. *Chem. Biol. Drug Des.* 2017, *89*, 420–427.

(14) Gao, M.; Duan, L.; Luo, J.; Zhang, L.; Lu, X.; Zhang, Y.; Zhang, Z.; Tu, Z.; Xu, Y.; Ren, X.; Ding, K. Discovery and optimization of 3-(2-(pyrazolo[1,5-a]pyrimidin-6-yl)ethynyl)benzamides as novel selective and orally bioavailable discoidin domain receptor 1 (ddr1) inhibitors. *J. Med. Chem.* 2013, *56*, 3281–3295.

(15) Wang, Z.; Zhang, Y.; Pinkas, D. M.; Fox, A. E.; Luo, J.; Huang, H.; Cui, S.; Xiang, Q.; Xu, T.; Xun, Q.; Zhu, D.; Tu, Z.; Ren, X.; Brekken, R. A.; Bullock, A. N.; Liang, G.; Ding, K.; Lu, X Design, synthesis, and biological evaluation of 3-(imidazo[1,2-a]pyrazin-3-ylethynyl)-4-isopropyl-n-(3-((4-methylpiperazin-1-yl)methyl)-5-(trifluoromethyl)phenyl)benzamide as a dual inhibitor of discoidin domain receptors 1 and 2. *J. Med. Chem.* 2018, *61*, 7977–7990.

(16) Wang, Z.; Bian, H.; Bartual, S. G.; Du, W.; Luo, J.; Zhao, H.; Zhang, S.; Mo, C.; Zhou, Y.; Xu, Y.; Tu, Z.; Ren, X.; Lu, X.; Brekken, R. A.; Yao, L.; Bullock, A. N.; Su, J.; Ding, K. Structure-based design of tetrahydroisoquinoline-7-carboxamides as selective discoidin domain receptor 1 (ddr1) inhibitors. *J. Med. Chem.* 2016, *59*, 5911–5916.

(17) Wang, Z.; Zhang, Y.; Bartual, S. G.; Luo, J.; Xu, T.; Du, W.; Xun, Q.; Tu, Z.; Brekken, R. A.; Ren, X.; Bullock, A. N.; Liang, G.; Lu, X.; Ding, K. Tetrahydroisoquinoline-7-carboxamide derivatives as new selective discoidin domain receptor 1 (ddr1) inhibitors. *ACS Med. Chem. Lett.* 2017, *8*, 327–332.

(18) Müller, S.; Ackloo, S.; Arrowsmith, C. H.; Bauser, M.; Baryza, J. L.; Blagg, J.; Böttcher, J.; Bountra, C.; Brown, P. J.; Bunnage, M. E.; Carter, A. J.; Damerell, D.; Dötsch, V.; Drewry, D. H.; Edwards, A. M.; Edwards, J.; Elkins, J. M.; Fischer, C.; Frye, S. V.; Gollner, A.; Grimshaw, C. E.; Ijzerman, A.; Hanke, T.; Hartung, I. V.; Hitchcock, S.; Howe, T.; Hughes, T. V.; Laufer, S.; Li, V. M.; Liras, S.; Marsden, B. D.; Matsui, H.; Mathias, J.; O'Hagan, R. C.; Owen, D. R.; Pande, V.; Rauh, D.; Rosenberg, S. H.; Roth, B. L.; Schneider, N. S.; Scholten, C.; Singh Saikatendu, K.; Simeonov, A.; Takizawa, M.; Tse, C.; Thompson, P. R.; Treiber, D. K.; Viana, A. Y.; Wells, C. I.; Willson, T. M.; Zuercher, W. J.; Knapp, S.; Mueller-Fahrnow, A. Donated chemical probes for open science. *eLife* 2018, *7*, No. e34311.

(19) Klein, R.; Silos-Santiago, I.; Smeyne, R. J.; Lira, S. A.; Brambilla, R.; Bryant, S.; Zhang, L.; Snider, W. D.; Barbacid, M. Disruption of the neurotrophin-3 receptor gene trkc eliminates la muscle afferents and results in abnormal movements. *Nature* 1994, *368*, 249–251.

(20) Smeyne, R. J.; Klein, R.; Schnapp, A.; Long, L. K.; Bryant, S.; Lewin, A.; Lira, S. A.; Barbacid, M. Severe sensory and sympathetic neuropathies in mice carrying a disrupted trk/ngf receptor gene. *Nature* 1994, *368*, 246–249.

(21) Schneider, H.; Szabo, E.; Machado, R. A. C.; Broggini-Tenzer, A.; Walter, A.; Lobell, M.; Heldmann, D.; Süssmeier, F.; Grünewald, S.; Weller, M. Novel tie-2 inhibitor bay-826 displays in vivo efficacy in experimental syngeneic murine glioma models. *J. Neurochem.* 2017, *140*, 170–182.

(22) Richter, H.; Satz, A. L.; Bedoucha, M.; Buettelmann, B.; Petersen, A. C.; Harmeier, A.; Hermosilla, R.; Hochstrasser, R.; Burger, D.; Gsell, B.; Gasser, R.; Huber, S.; Hug, M. N.; Kocer, B.; Kuhn, B.; Ritter, M.; Rudolph, M. G.; Weibel, F.; Molina-David, J.; Kim, J.-J.; Santos, J. V.;

Stihle, M.; Georges, G. J.; Bonfil, R. D.; Fridman, R.; Uhles, S.; Moll, S.; Faul, C.; Fornoni, A.; Prunotto, M. DNA-encoded library-derived ddr1 inhibitor prevents fibrosis and renal function loss in a genetic mouse model of alport syndrome. *ACS Chem. Biol.* 2019, *14*, 37–49.

(23) Petersen, L. K.; Blakskjær, P.; Chaikuad, A.; Christensen, A. B.; Dietvorst, J.; Holmkvist, J.; Knapp, S.; Kořínek, M.; Larsen, L. K.; Pedersen, A. E.; Röhm, S.; Sløk, F. A.; Hansen, N. J. V. Novel p38α map kinase inhibitors identified from yocto reactor dna-encoded small molecule library. *MedChemComm* 2016, *7*, 1332.

(24) Röhm, S.; Schröder, M.; Dwyer, J. E.; Widdowson, C. S.; Chaikuad, A.; Berger, B.-T.; Joerger, A. C.; Krämer, A.; Harbig, J.; Dauch, D.; Kudolo, M.; Laufer, S.; Bagley, M. C.; Knapp, S. Selective targeting of the αc- and dfg-out pocket in p38 mapk. *Eur. J. Med. Chem.* 2020, *208*, 112721.

(25) Röhm, S.; Berger, B.-T.; Schröder, M.; Chaikuad, A.; Winkel, R.; Hekking, K. F. W.; Benninghof, J. J. C.; Müller, G.; Tesch, R.; Kudolo, M.; Forster, M.; Laufer, S.; Knapp, S. Fast iterative synthetic approach toward identification of novel highly selective p38 map kinase inhibitors. *J. Med. Chem.* 2019, *62*, 10757–10782.

(26) Ahn, Y. M.; Clare, M.; Ensinger, C. L.; Hood, M. M.; Lord, J. W.; Lu, W.-P.; Miller, D. F.; Patt, W. C.; Smith, B. D.; Vogeti, L.; Kaufman, M. D.; Petillo, P. A.; Wise, S. C.; Abendroth, J.; Chun, L.; Clark, R.; Feese, M.; Kim, H.; Stewart, L.; Flynn, D. L. Switch control pocket inhibitors of p38-map kinase. Durable type ii inhibitors that do not require binding into the canonical atp hinge region. *Bioorg. Med. Chem. Lett.* 2010, *20*, 5793.

(27) Zhu, D.; Huang, H.; Pinkas, D. M.; Luo, J.; Ganguly, D.; Fox, A. E.; Arner, E.; Xiang, Q.; Tu, Z.-C.; Bullock, A. N.; Brekken, R. A.; Ding, K.; Lu, X. 2-amino-2,3-dihydro-1h-indene-5-carboxamide-based discoidin domain receptor 1 (ddr1) inhibitors: Design, synthesis, and in vivo antipancreatic cancer efficacy. *J. Med. Chem.* 2019, *62*, 7431–7444.

(28) Ruiz-Castro, P. A.; Shaw, D.; Jarai, G. Discoidin domain receptor signaling and pharmacological inhibitors. In *Discoidin Domain Receptors in Health and Disease*; Fridman, R., Huang, P. H., Eds.; Springer New York: New York, NY, 2016; pp 217–238.

(29) Matsuyama, W.; Wang, L.; Farrar, W. L.; Faure, M.; Yoshimura, T. Activation of discoidin domain receptor 1 isoform b with collagen up-regulates chemokine production in human macrophages: Role of p38 mitogen-activated protein kinase and nf-κb. *J. Immunol.* 2004, *172*, 2332–2340.

(30) Fedorov, O.; Niesen, F. H.; Knapp, S. Kinase inhibitor selectivity profiling using differential scanning fluorimetry. In *Kinase Inhibitors: Methods and Protocols*; Kuster, B., Ed.; Humana Press: Totowa, NJ, 2012; pp 109–118.

(31) Vasta, J. D.; Corona, C. R.; Wilkinson, J.; Zimprich, C. A.; Hartnett, J. R.; Ingold, M. R.; Zimmerman, K.; Machleidt, T.; Kirkland, T. A.; Huwiler, K. G.; Ohana, R. F.; Slater, M.; Otto, P.; Cong, M.; Wells, C. I.; Berger, B.-T.; Hanke, T.; Glas, C.; Ding, K.; Drewry, D. H.; Huber, K. V. M.; Willson, T. M.; Knapp, S.; Müller, S.; Meisenheimer, P. L.; Fan, F.; Wood, K. V.; Robers, M. B. Quantitative, wide-spectrum kinase profiling in live cells for assessing the effect of cellular atp on target engagement. *Cell Chem. Biol.* 2018, *25*, 206.

(32) Machleidt, T.; Woodroofe, C. C.; Schwinn, M. K.; Méndez, J.; Robers, M. B.; Zimmerman, K.; Otto, P.; Daniels, D. L.; Kirkland, T. A.; Wood, K. V. Nanobret—a novel bret platform for the analysis of protein–protein interactions. *ACS Chem. Biol.* 2015, *10*, 1797.

(33) De Nicola, G. F.; Martin, E. D.; Chaikuad, A.; Bassi, R.; Clark, J.; Martino, L.; Verma, S.; Sicard, P.; Tata, R.; Atkinson, R. A.; Knapp, S.; Conte, M. R.; Marber, M. S. Mechanism and consequence of the autoactivation of p38α mitogen-activated protein kinase promoted by tab1. *Nat. Struct. Mol. Biol.* 2013, *20*, 1182–1190.

(34) Winter, G.; Lobley, C. M. C.; Prince, S. M. Decision making in xia2. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2013, *69*, 1260–1273.

(35) Kabsch, W. Xds. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2010, *66*, 125.

(36) Evans, P. R.; Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2013, *69*, 1204.

(37) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* 2007, 40, 658.

(38) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2010, 66, 486.

(39) Kovalevskiy, O.; Nicholls, R. A.; Long, F.; Carlon, A.; Murshudov, G. N. Overview of refinement procedures within refmac5: Utilizing data from different sources. *Acta Crystallogr., Sect. D: Struct. Biol.* 2018, 74, 215−227.

(40) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. Phenix: A comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2010, 66, 213−221.

(41) Howarth, A.; Schröder, M.; Montenegro, R. C.; Drewry, D. H.; Sailem, H.; Millar, V.; Müller, S.; Ebner, D. V. High via-a flexible live-cell high-content screening pipeline to assess cellular toxicity. *SLAS Discovery* 2020, 25, 801−811.

(42) Bauer, S. M.; Kubiak, J. M.; Rothbauer, U.; Laufer, S. From enzyme to whole blood: Sequential screening procedure for identification and evaluation of p38 mapk inhibitors. *Methods Mol. Biol.* 2015, 1360, 123−148.

Appendix H: Discovery of 3-Amino-1h-Pyrazole-Based Kinase Inhibitors to Illuminate the Understudied Pctaire Family

Reprinted with permission from: Amrhein, Jennifer Alisa, Lena Marie Berger, Amelie Menge, Andreas Krämer, Lewis Elson, Tuomas Tolvanen, Daniel Martinez-Molina, Astrid Kaiser, Manfred Schubert-Zsilavecz, Susanne Müller, Stefan Knapp, and Thomas Hanke. "Discovery of 3-Amino-1h-Pyrazole-Based Kinase Inhibitors to Illuminate the Understudied Pctaire Family." International Journal of Molecular Sciences 23, no. 23 (2022): 14834. https://www.mdpi.com/1422-0067/23/23/14834. Copyright 2022 by MDPI.

Notice that further permission related to the material excerpted should be directed to the publisher.

Contribution Menge, A.: Conducting fluorescent ubiquitination-based cell cycle reporter (FUCCI) assay, contribution to study design, data evaluation, figure preparation, manuscript writing and editing.

International Journal of
Molecular Sciences

MDPI

Article

Discovery of 3-Amino-1*H*-pyrazole-Based Kinase Inhibitors to Illuminate the Understudied PCTAIRE Family

Jennifer Alisa Amrhein [1,2], Lena Marie Berger [1,2], Amelie Tjaden [1,2], Andreas Krämer [1,2,3], Lewis Elson [1,2], Tuomas Tolvanen [4], Daniel Martinez-Molina [5], Astrid Kaiser [1], Manfred Schubert-Zsilavecz [1], Susanne Müller [1,2], Stefan Knapp [1,2,3,*] and Thomas Hanke [1,2,*]

[1] Institute of Pharmaceutical Chemistry, Goethe-University Frankfurt, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany
[2] Structural Genomics Consortium, Buchmann Institute for Life Sciences, Goethe-University Frankfurt, Max-von-Laue-Str. 15, 60438 Frankfurt am Main, Germany
[3] German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), DKTK Site Frankfurt-Mainz, 69120 Heidelberg, Germany
[4] Division of Rheumatology, Department of Medicine Solna, Karolinska University Hospital and Karolinska Institute, Solnavägen 1, 17177 Solna, Sweden
[5] Pelago Bioscience AB, Scheeles Väg 1, 17165 Solna, Sweden
* Correspondence: knapp@pharmchem.uni-frankfurt.de (S.K.); hanke@pharmchem.uni-frankfurt.de (T.H.)

Abstract: The PCTAIRE subfamily belongs to the CDK (cyclin-dependent kinase) family and represents an understudied class of kinases of the dark kinome. They exhibit a highly conserved binding pocket and are activated by cyclin Y binding. CDK16 is targeted to the plasma membrane after binding to N-myristoylated cyclin Y and is highly expressed in post-mitotic tissues, such as the brain and testis. Dysregulation is associated with several diseases, including breast, prostate, and cervical cancer. Here, we used the N-(1H-pyrazol-3-yl)pyrimidin-4-amine moiety from the promiscuous inhibitor 1 to target CDK16, by varying different residues. Further optimization steps led to 43d, which exhibited high cellular potency for CDK16 (EC_{50} = 33 nM) and the other members of the PCTAIRE and PFTAIRE family with 20–120 nM and 50–180 nM, respectively. A DSF screen against a representative panel of approximately 100 kinases exhibited a selective inhibition over the other kinases. In a viability assessment, 43d decreased the cell count in a dose-dependent manner. A FUCCI cell cycle assay revealed a G2/M phase cell cycle arrest at all tested concentrations for 43d, caused by inhibition of CDK16.

Keywords: CDK16; cell cycle; FUCCI; kinase; kinase inhibitor; NanoBRET; synthesis

1. Introduction

Cyclin-dependent kinases are serine-threonine kinases activated by activation loop phosphorylation and cyclins. The CDK family in eukaryotic cells consists of 21 serine/threonine kinases that are characterized by conserved structural motifs and sequence similarity (Figure 1A) [1]. They are characterized by a largely conserved ATP binding pocket, a PSTAIRE-like binding domain, and an activating T-loop motif. Activation of the kinase through cyclin binding shifts the T-loop to expose the substrate-binding site and primes the phospho-transfer reaction by aligning specific residues within the active site [2,3]. CDKs have key roles in regulating the cell cycle but they also perform diverse other cellular functions, such as regulation of mRNA processing and transcription [4–6]. Many CDKs have evolved as major drug targets which has led to recent approvals of CDK4/6 inhibitors for the treatment of cancer [7–9].
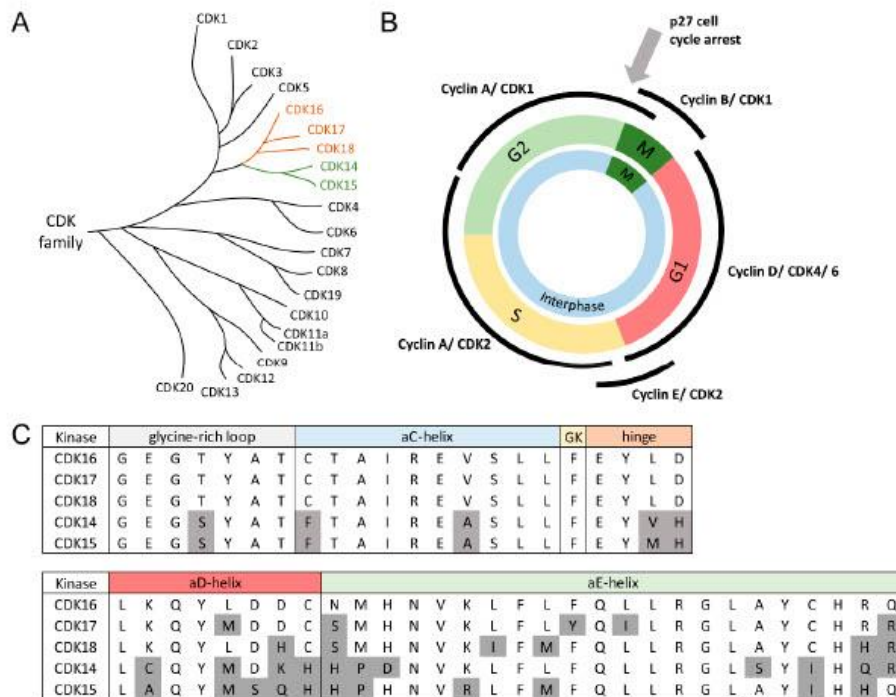
Figure 1. (**A**) Phylogenetic tree of the CDK family. The PFTAIRE subfamily is highlighted in green, the PCTAIRE subfamily in orange (Adapted from Wells et. al.,) [1]. (**B**) Overview of the cell cycle phases and the involvement of CDK's 1,2,4 and 6 in the transitions through the different phases (Adapted from Suryadinata et.al.,) [10]. Dysregulation of p27 through CDK16 led to a G2/M phase cell cycle arrest. (**C**) Alignment of key residues of the ATP binding pocket of CDK14–18. Differences in comparison to CDK16 are highlighted in grey.

However, CDK14–18, also called "TAIRE" kinases based on a conserved sequence motif, remain poorly studied and no selective inhibitors have been developed so far, despite interesting links to disease development. CDK14, for instance, is implicated in the regulation of Wnt signaling by forming a CDK14-cyclin Y complex at the plasma membrane that phosphorylates the Wnt co-receptor Lrp6 [11]. CDK14 together with CDK15 forms a small subfamily also referred to as PFTK1 and PFTK2, based on the sequence of the cyclin-binding domain (PFTAIRE). Additionally, CDK15 regulates the Wnt pathway and β-catenin/MEK–ERK signaling by phosphorylation of PAK4, promoting cell proliferation and migration. CDK15 activation has therefore been linked to colorectal cancer progression [12]. CDK16–18 (PCTK1–3) form the PCTAIRE subfamily, which shares high sequence similarity with small differences within their N- and C-terminal domains [13]. CDK17 plays a role in regulating the glycerophospholipid metabolism and it shows high expression levels in Alzheimer's disease where it promotes neurodegeneration [14]. Dysfunction of CDK18 has been associated with Alzheimer's disease and other diseases, such as neurological disorders, cerebral ischemia, and cancer [15]. CDK16 is activated by cyclin Y at the plasma membrane and participates in many different pathways including Wnt-dependent signalling or signal transduction in the primary cilium [16]. In contrast to other CDKs, binding of cyclin Y to

CDK16 requires a part of an N-terminal extension in addition to the kinase domain. This additional region contains serine 153, which can be phosphorylated. Phosphorylation at this site prevents the formation of the CDK16-cyclin Y complex preventing CDK16 activation [17]. CDK16 is highly expressed in postmitotic tissues such as the testis and the brain. Furthermore, CDK16 is highly expressed in postmeiotic spermatids, whereas it is absent in mature spermatozoa [18]. In the brain, CDK16 can be found in Purkinje and pyramidal cells of the hippocampus and in the neocortex and plays a role in regulation of intracellular vesicles and neurite outgrowth [19,20]. It is also involved in other processes such as glucose homeostasis [21], vesicle trafficking [22], and muscle differentiation [23]. Dysregulation of CDK16 has been described in many cancers such as breast, prostate, cervical cancers, and melanomas [24–26]. For example, CDK16 phosphorylates the tumor suppressor p27 at Ser10, which promotes ubiquitination, thereby attenuating apoptosis [27]. p27 regulates the cell cycle by inhibiting CDK4 and CDK6/cyclin complexes, as well as influencing the motility and apoptosis of cells [28]. Downregulation of p27 is associated with high-grade tumors and a poor survival prognosis [29,30]. Yanagi et. al. demonstrated that CDK16 knockout induces G2/M cell cycle arrest, apoptosis, and accumulation of p27 in cutaneous squamous cell carcinoma cells (SCC) (Figure 1B) [28]. In 2018, Wang et. al. found increased expression of CDK16 in NSCLC tumor tissues. Therefore, the oncogenic role of CDK16 is attributed to the p27 ubiquitination triggered by the phosphorylation [31]. Degradation of the tumor suppressor p53 via ubiquitination is also attributed to CDK16. It has been shown that CDK16 phosphorylates p53 at Ser315 and enhances the transcriptional activity which promotes radioresistance in lung cancer [32]. In addition, Dohmen et. al. identified the CDK16/cyclin Y complex as a substrate of AMPK in macroautophagy. Cyclin Y is phosphorylated by AMPK at Ser326, which promotes the activity of CDK16 by forming the CDK16-cyclin Y complex and a subsequent autophosphorylation at Ser336. These phosphorylation events induce autophagy in a ULK1 and Beclin1-dependent manner [33,34]. Despite their biological importance, CDK14–18 represents an understudied subclass of CDKs, whereas CDK1, 2, and 4, in particular, are well characterized [35]. This highlights the need to develop new inhibitors to gain a better understanding of the functions of these understudied kinases.

CDK inhibitors represent an important class of anti-tumor therapeutics. The development of CDK inhibitors extends to the 1990s. The first ones were pan-CDK inhibitors, such as Flavopiridol or Roscovitine, which affected cell proliferation. Due to the low selectivity and high toxicity of the so-called first generation of pan-CDK inhibitors, most of them failed in clinical trials. The second generation of pan-CDK inhibitors led to compounds such as dinaciclib or roniciclib with an improved selectivity profile and consequently fewer side effects. To date, approximately 40 pan-CDK inhibitors are in diverse stages of research or different phases of clinical trials and are used for the treatment of various tumors, such as leukaemia, melanoma, or breast cancer. Nevertheless, the development of specific CDK inhibitors represents an important research area in the medicinal chemistry. Palbociclib, ribociclib, and abemaciclib are FDA-approved CDK4/6 inhibitors for the treatment of HR+ advanced breast cancer. They interfere with the cell cycle by blocking the phosphorylation of Rb protein, resulting in cell cycle arrest from G1 to S phase. ICEC0942 is a phase I/II CDK7 inhibitor for the treatment of breast or prostate cancer. AZD-4573 inhibits CDK9, leading to a down-regulation of oncogenic genes such as MCL-1. Preclinical studies revealed anti-tumor effects in hematologic malignancies. The covalent CDK12/13 inhibitor THZ531 exhibits down-regulation effects of DNA damage response genes and is used together with sorafenib for the treatment of hepatocellular carcinoma [36,37]. To date, only a covalent pan-TAIRE inhibitor has been developed [38]. However, some PARP inhibitors have been recently described targeting kinases including CDK15 [39] and the structure of CDK16 with non-selective kinases inhibitors that may serve as starting points for chemical probe development has been disclosed [40]. The development of selective inhibitors for the PCTAIRE subfamily is still in the early stages and to date, no selective inhibitors are published. The high similarity of the kinase domain within the PCTAIRE subfamily with 73% identity makes it challenging to develop selective kinase inhibitors (Figure 1C) [26].

In 2008, Statsuk et. al. published the promiscuous kinase inhibitors **1**, **2** and **3** (Figure 2A). A kinome-wide screen against 359 wild-type kinases emphasized their promiscuous behaviour by targeting 337 and 317 kinases for **1** and **3**, respectively. The selectivity profile of **1** at a screening concentration of 1 μM is shown in Figure 2B. Additionally, the K_D values of **1** against 40 different kinases have been reported. CDK2, CDK5, and JNK3 were the most potently inhibited kinases in this set with 4.6, 27.6, and 26.1 nM, respectively [41]. The crystal structures of all compounds (**1–3**) were each determined in complex with different kinases such as VRK1 with **1**, STK17B with **2**, and c-Src with **3** (Figure 2C–E). These different crystal structures indicated mainly two different binding modes. While for **1** the pyrimidine is facing towards the hydrophobic pocket, in **2** and **3** the molecule is reversed and the pyrimidine or quinazoline moiety targets the front pocket. The flexibility of the molecule in the binding pocket could be an explanation for the various targets and the promiscuous behaviour. The high on-target potency within the CDK family, particularly the PCTAIRE family [41], combined with the known crystal structures of the various derivatives makes the 3-aminopyrazole moiety an excellent starting point for the development of a structure activity relationship (SAR) study to obtain a selective kinase inhibitor for exploring the dark kinase CDK16.
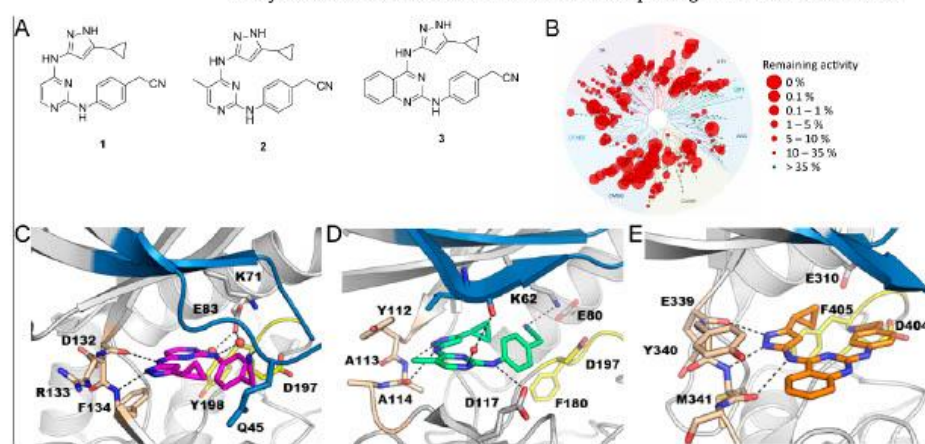
Figure 2. (A). Chemical structures of three previous published promiscuous kinase inhibitors **1**, **2** and **3**. (B). Selectivity profile of **1** against 359 recombinant human protein kinases at a screening concentration of 1 μM, using the KINOMEscan assay platform [41]. (C). Crystal structure of VRK1 in complex with **1** (PDB: 3OP5). (D). Crystal structure of STK17B in complex with **2** (PDB: 3LM0). (E). Crystal structure of c-Src in complex with **3** (PDB: 3F6X). The hinge region is highlighted in beige, the P-loop in blue, and the DFG motif (in VRK1 DYG motif) in yellow.

The pyrazole hinge binding moiety represents a privileged scaffold in the medicinal chemistry for the development of kinase inhibitors. Previous publications underline the importance, as well as the anti-proliferative and anti-cancer potential of pyrazole-based molecules. They play a crucial role in the treatment of various diseases and cancer types, such as breast cancer, lymphoma, cervical cancer, or inflammation disorders [42,43].

Here, the N^4-(1H-pyrazol-3-yl)pyrimidine-2,4-diamine core of the promiscuous inhibitors (**1–3**) was chosen as pharmacophore for the structure-based modification on this scaffold (Figure 3). The N-(1H-pyrazol-3-yl)pyrimidin-4-amine core can be also found in the aurora kinase inhibitor tozasertib. It harbors a methyl moiety at the pyrazole, an additional residue at the C6 position of the pyrimidine, and an aromatic linker that is connected through a sulphur atom. It exhibited a better selectivity profile than the promiscuous

inhibitor **1**, showed no affinity for CDK2 and CDK5, however, targeted CDK16 with a K_D value of 160 nM [44,45]. Those results underlined the hypothesis to gain selectivity by varying the different substituents at the pyrazole and the pyrimidine. Three different positions were selected to determine their influence on selectivity and potency against CDK16. Therefore, the head group on the pyrazole was varied. Different smaller and bulkier alkyl groups, ester, and amide groups were tested. Furthermore, the effect of an additional residue at the pyrimidine was explored. Based on the lead structure, a methyl-, a chloro-, and a phenyl-residue were used and different linkers were selected for the third position in our SAR analysis. Derived from the nitril of **1**, Boc-protected amines in different positions of the aromatic ring were explored first. In addition, the position of the aromatic ring was varied, alkyl linkers were used and the linker length was changed. The influence of the Boc group was also tested by replacing it with a smaller amide and free amines.
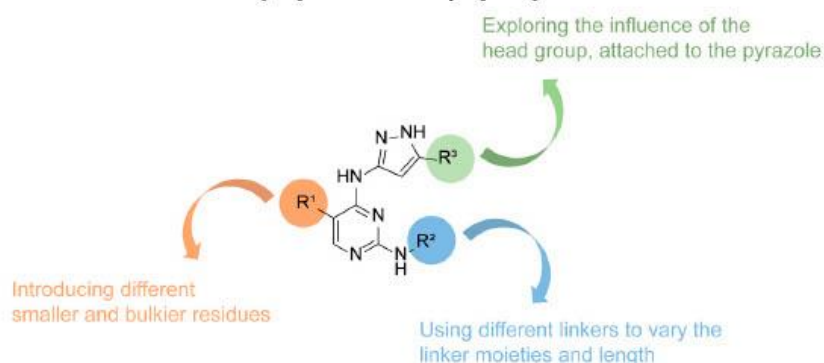
Figure 3. Summary of modifications, using **1–3** as the lead structure.

2. Results

2.1. Synthesis of 3-Aminopyrazole-Based Kinase Inhibitors

Based on the scaffold of **1**, the first set of molecules was synthesized. Therefore, 5-cyclopropyl-1*H*-pyrazole-3-amine (**4**) and a pyrimidine derivative (**5–7**) were used for a nucleophilic substitution under basic conditions, whereby different substituents such as a hydrogen, a methyl, and a chloro residue at position 5 of the pyrimidine were used. The yields were in the range of 35–89%. Various linkers were attached to the precursor via a second nucleophilic substitution under basic conditions under microwave irradiation or with a catalytic amount of 1M HCl. For this purpose, different aniline derivatives, benzylamine derivatives, and aliphatic linkers were used to obtain the final compounds **11a–f**, **12a–e**, and **13a, c**. The aniline derivatives could be obtained in moderate to good yields, ranging from 16 to 90%. However, the aliphatic linkers required quite harsh reaction conditions with long reaction times under microwave irradiation and ended in yields of 5–52%. To determine the influence of the Boc group in comparison to a smaller residue, the protecting group of **11e** was cleaved with TFA in DCM. A smaller residue was introduced via an amide coupling, using HATU and acetic acid under basic conditions to obtain **15** with a 30% yield (Scheme 1).
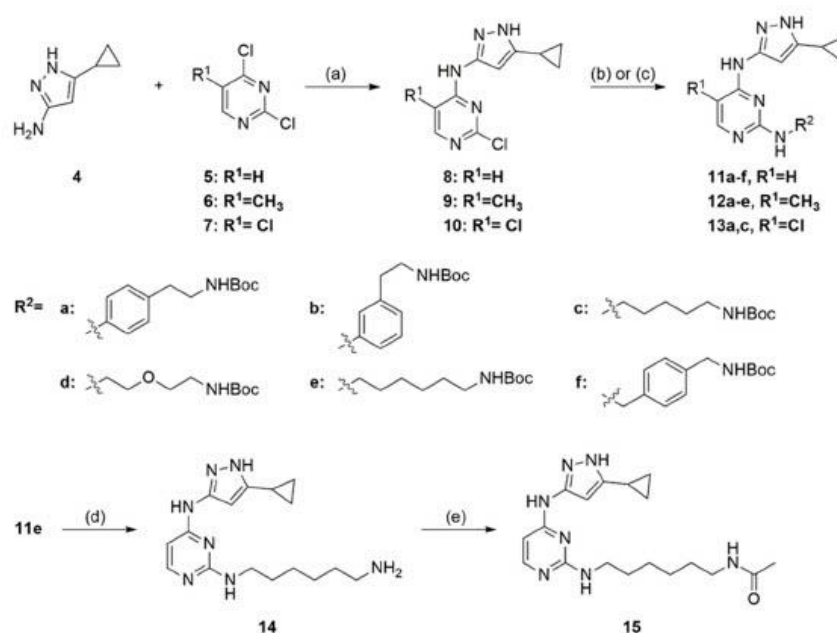
Scheme 1. Synthesis of the first series of derivatives **11a–f, 12a–e, 13a, c,** and **15**. a. Reagents and conditions: (a) TEA and isopropanol, 48 h, 55–80 °C; (b) TEA and ethanol, microwave, 5–10 h, 80–90 °C; (c) HCl and ethanol, reflux, 4–18h; (d) TEA and DCM, 0 °C–rt, oN; (e) Acetic acid, HATU, DIPEA and DMF, rt, oN.

For the second series, we replaced the cyclopropyl group on the pyrazole with a methyl ester. Again, we used different substituents such as a hydrogen or a chloro residue at position 5 of the pyrimidine, or a quinazoline moiety instead of the pyrimidine and different linkers to investigate their influence. Compounds **18–20** were received in yields between 16–84%. The final compounds **21a–i, 22a–e,** and **23a–c** were synthesized as described above in a two-step synthesis route. The aniline derivatives could be obtained in moderate yields between 49–72% for **21** and **23**. However, a chloro residue at position 5 of the pyrimidine (**22**) decreased the yields dramatically to 5–9%. The yields for the introduction of the aliphatic linkers ranged from 17 to 64%. The Boc group was cleaved for **21a** and **21i** to determine the influence of the primary amines. **24** and **25** were received with 54% and 36%, respectively (Scheme 2). The moiety on the pyrazole was exchanged through various residues like a methyl, isopropyl, *tert*-butyl, methyl amide, isopropyl ester, or *tert*-butyl ester group in the third series. The alkyl groups (**32–34**) and the ester moieties (**36–37**) could be obtained in moderate yields ranging from 35–84%. However, the amide (**35**) was only achieved with 14%. One to three different linkers were attached for each of these functional groups to obtain the molecules **38–43**. For the second nucleophilic substitution, compounds **38–41** were gained with yields between 22 and 79%. The compounds harboring an ester moiety (**42–43**) reduced the turnover rate to 4–12% (Scheme 3).
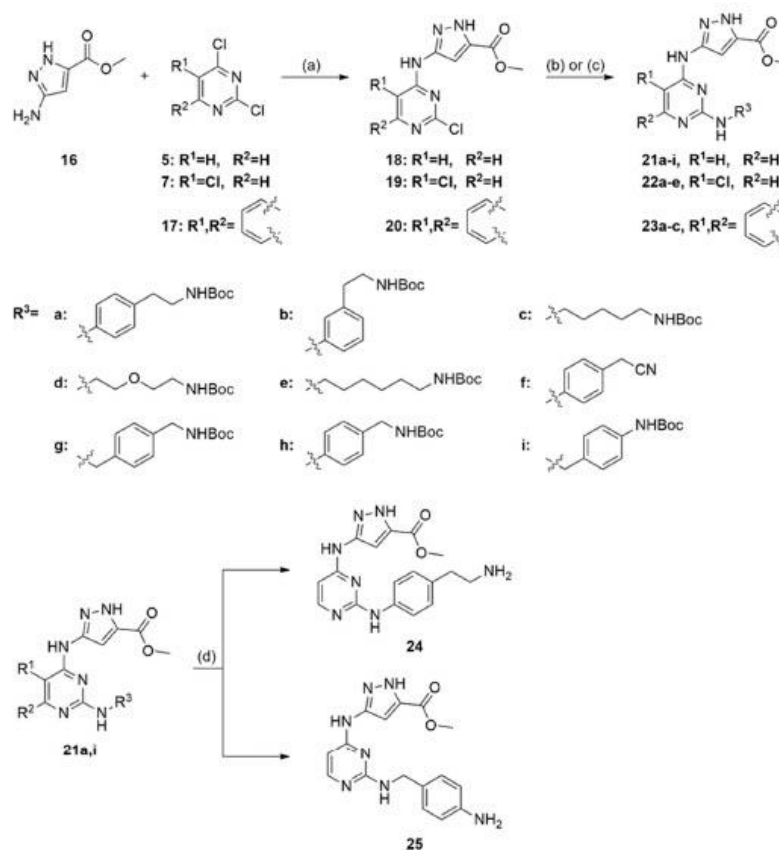
Scheme 2. Synthesis of the second series of derivatives **21a–i, 22a–e, 23a–c,** and **24–25**. a. Reagents and conditions: (a) TEA and isopropanol, 24–72 h, 50–60 °C; (b) TEA and methanol or ethanol, microwave, 3–8 h, 80–120 °C; (c) HCl and ethanol, 70 °C–reflux, 18 h; (d) TFA and DCM, 0 °C–rt, oN.

2.2. Structure-Activity Relationship of 3-Aminopyrazole-Based Molecules

To investigate the selectivity profile of the different series, a differential scanning fluorimetry (DSF) assay was used [46]. This assay offers a rapid and sensitive screening method that determines the denaturation temperature of a protein in absence and presence of a compound. This assay format measures the fluorescence of a dye, which depends on the folding state of the protein. The difference between the melting temperature (ΔTm) in the absence and presence of the compound represents the binding strength to the protein. We used an in-house panel of 89–105 kinases to screen all synthesized inhibitors (Tables S1–S3).

The lead structure **1** was resynthesized, based on the synthetic route of Statsuk et. al. [41]. Using the DSF assay format, we identified several molecules, which showed a strong stabilization of CDK16 (Figure 4), including targets that were reported to be strongly inhibited by lead structure **1**. For example, CDK2 and JNK3, as well as GSK3B, which is a potential off-target of the newly synthesized compounds are shown together with the

number of kinases with ΔTm shifts >5 °C, providing an indication of the overall selectivity. The compounds of the first series that lacked substitution at the pyrimidine (**11a–11f**) exhibited a high stabilization of CDK16. Especially **11c** showed a comparable ΔTm shift with 9.4 °C in comparison to **1** (10.3 °C). The stabilization of CDK2 was also quite high; however, the shifts of JNK3 were negligible. The introduction of a methyl or chloro residue at the pyrimidine was not tolerated. An additional methyl group (**12a–e**) reduced the ΔTm shifts by 2 °C on average in comparison to the corresponding molecule without the methyl group. An even stronger reduction of ΔTm shifts was observed by the introduction of a bulkier chloro residue (**13a**, **13c**) showing ΔTm shifts of only 3.2 °C. Additionally, the stabilization of CDK2 was decreased by introduction of the residues at R^1; however, a chloro residue was tolerated screening JNK3 and the presence of this R^1 group increased ΔTm shifts to 4.9–7.0 °C. The replacement of the Boc group to a smaller amide (**15**) had no influence on ΔTm shifts.
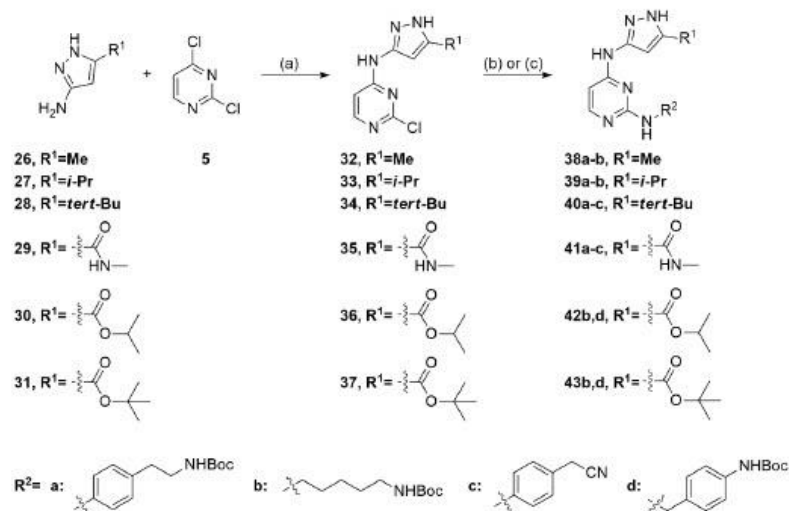
Scheme 3. Synthesis of the third series of derivatives **38–39a–b**, **40–41a–c**, and **42–43b, d**. a. Reagents and conditions: (a) TEA and isopropanol, 18–120 h, 60 °C; (b) TEA and ethanol or isopropanol or *tert*-butanol, microwave, 8 h, 90–100 °C; (c) HCl and ethanol, 70 °C–reflux, 18 h.

The lead structure **1** was highly promiscuous as indicated by stabilizing 60 kinases with ΔTm shifts >5 °C. The compounds with the same precursor but an exchanged linker moiety **11a–f** and **15** were also nonselective with 12–32 significantly stabilized kinases. Additionally, the introduction of an additional residue, such as a methyl or chloro residue at the pyrimidine (**12a–12e**, **13a**, and **13c**), did not lead to an improved selectivity with 16–28 shifts >5 °C. The best selectivity within this series was observed for compounds **11d** and **12d** with the *tert*-butyl (2-(aminomethoxy)ethyl)carbamate linker, whereas the replacement from **11e** with the smaller amide in **15** did not impact selectivity.

The introduction of the methyl ester at the pyrazole (**21a–i**) led to an overall lower stabilization of CDK16 in comparison to the lead structure **1**; however, **21c** and **21e** exhibited still similar ΔTm shifts compared to staurosporine (9.1 °C), which has a reported K$_D$ value of 24 nM [47]. Additionally, **21i** showed significant ΔTm stabilization (6.5 °C). Surprisingly, the 4-aminobenzonitrile linker (**21f**) exhibited a high stabilization of GSK3B and JNK3, while CDK16 and CDK2 showed with 5.3 and 7.8 °C moderate ΔTm values. Additionally, here, modifications at R^1 were not tolerated by CDK16. All molecules with the bulkier chloro

residue at this position (22a–22e) or the exchange of the pyrimidine to a quinazoline moiety (23a–23c) showed no stabilization in the DSF assay. Furthermore, in this second series we were interested in the overall impact of the Boc group and cleaved these compounds to the primary amine in compounds 24 and 25, which led to a decreased stabilization of CDK16.
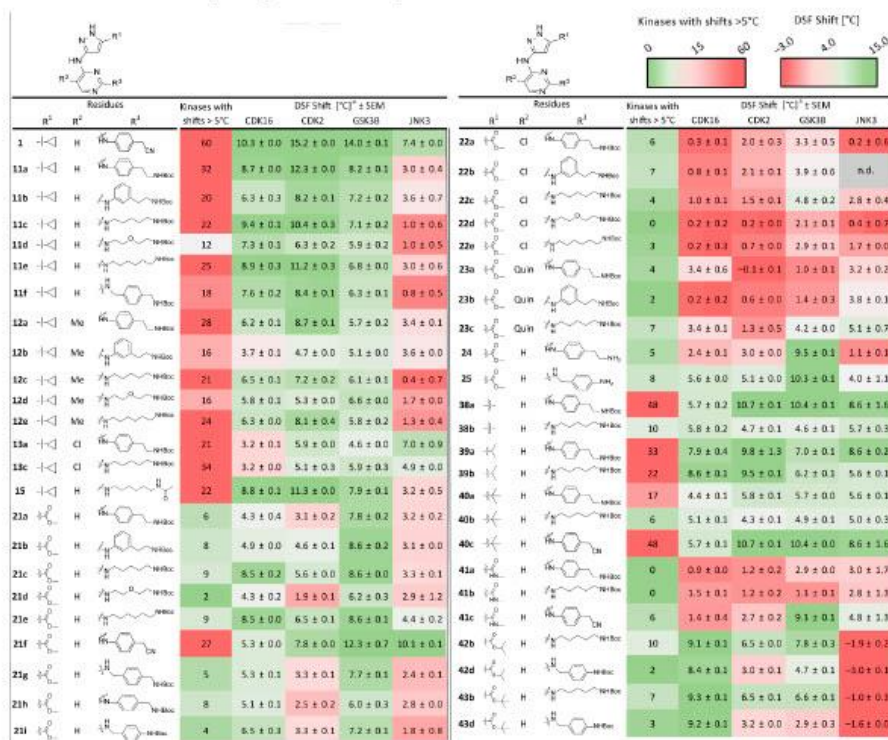
Figure 4. Heatmap of the DSF results of the N-(1H-pyrazole-3-yl)pyrimidine-4-amine-based inhibitors against CDK16, CDK2, GSK3B, and JNK3 (red means a low stabilization of the desired kinase, green a high stabilization). The number of kinases with shifts >5 °C provided an indication of the selectivity profile of the newly synthesized compounds (red indicates a high number of stabilized kinases and therefore a promiscuous behaviour, green a low number of stabilized kinases). [a] DSF measurements were performed in duplicates (±SEM).

Gratifyingly, the second series showed in general a better selectivity profile and compounds of this series stabilized only between 0–9 kinases with a ΔTm shift > 5 °C. In analogy to the first series, substitution at R^1 (22a–22e, 23a–23c) had no influence compared to compounds that had a hydrogen at this position (21a–21i). Furthermore, the $tert$-butyl (2-(aminomethoxy)ethyl)-carbamate linker (21d, 22d) improved selectivity and led also to a completely inactive compound 22d. Interestingly, 21f revealed strong ΔTm shifts of 27 kinases. It seemed that not only the cyclopropyl, but also the 4-aminobenzonitrile linker increase inhibitor promiscuity. The cleavage of the Boc group led to no significant influence on compound selectivity. 24 exhibited the same amount of potential targets, whereby the amount was even increased for 25.

The introduction of a methyl, *iso*-propyl, or *tert*-butyl group in the third series led to moderate (38a–38b, 40a–40c) to good (39a–39b) stabilization of CDK16. Nevertheless, these modifications exhibited also high ΔTm shifts for CDK2, GSK3B, and JNK3. The replacement to a methyl amide moiety led surprisingly to inactive compounds with a ΔTm stabilization of 0.9–1.5 °C for CDK16. Only the 4-aminobenzonitrile linker (40c) resulted in ΔTm shift of GSK3B of 9.3 °C. Next, an *iso*-propyl- (42b, 42d) and *tert*-butyl (43b, 43d) ester were introduced at the pyrazole. These moieties can be found in various already approved drugs or those who are in clinical trials indicating that these groups might have valuable pharmacokinetic properties [48,49]. This led to increased ΔTm shifts between 8.4–9.3 °C for CDK16, while the stabilization for CDK2 and GSK3B was in a moderate range with 3.0–6.5 °C and 2.9–7.8 °C, respectively. JNK3 did not tolerate these variations with DSF shifts below or around 0 °C.

The introduction of a methyl, *iso*-propyl, or *tert*-butyl group again led to an increased number of ΔTm shifts >5 °C. The number of hits was between 6 and 48 with the trend that the bulkier *tert*-butyl group (40a–40c) was more selective than the smaller methyl- or *iso*-propyl groups (38a–b, 39a–b). Remarkably, a methyl amide group was not tolerated in this position. 41a and 41b were completely inactive, only 41c showed residual shifts due to the promiscuous 4-aminobenzonitrile linker. The replacement of the methyl ester to a more stable *iso*-propyl or *tert*-butyl ester decreased the number of potential off-targets. Only 1–2 kinases beside CDK16 exhibited a DSF shift >5 °C for 42d and 43d (Figure 5). However, these shifts were comparatively small compared to the staurosporine reference, suggesting that these compounds did not bind strongly to these kinases. 42b and 43b exhibited 7–10 ΔTm shifts >5 °C, which is in the same range as the corresponding compound 21c which has a methyl ester at this position.

Figure 5. (A). Graphic representation of the selectivity data of 1 preserved from the DSF assay, illustrated with Cell Signaling Technolo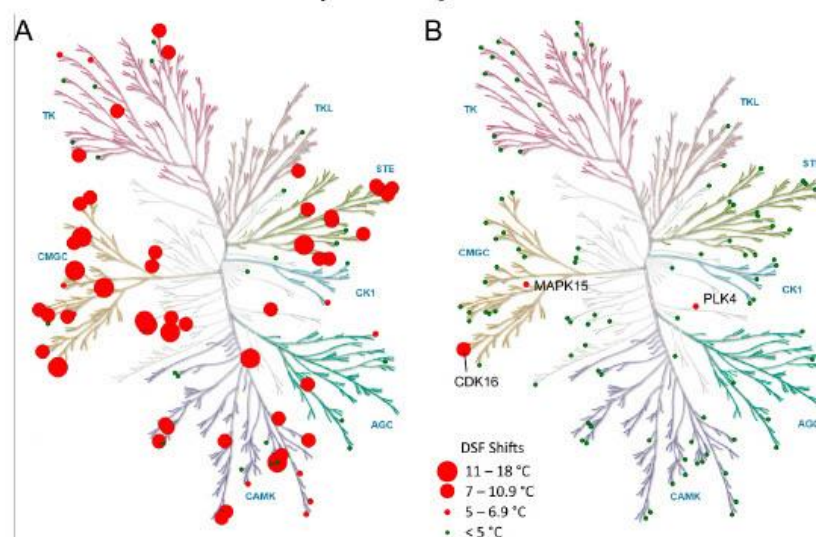gy. (B). Graphic representation of the selectivity data of 43d preserved from the DSF assay, illustrated with Cell Signaling Technology.

In conclusion, the introduction of the cyclopropyl group achieved the highest ΔTm shifts for CDK16 in the DSF assay; however, these compounds had a high number of other potential off-targets. A bulkier substituent at the pyrimidine, as well as a methyl

amide group at the pyrazole, were not tolerated by CDK16. The most favorable combination of CDK16 stabilization and potential off-targets was observed for **21c**, **21e**, **21g**, **21h**, and **21i** which belonged to the second series, and **42d** and **43d** carrying an *iso*-propyl or *tert*-butyl ester.

2.3. 3-Aminopyrazole-Based Molecules Selectively Stabilized CDK16 in Cells

To verify whether the substances also lead to stabilization of proteins in the cellular system, proteome-wide CETSA experiments were performed. Therefore, lead structure **1** and some compounds from the second series (**21b**, **21d**, **21e**, **21h**, and **21i**) were selected for this study (Figure 6, Figure S1). While the statistical significance is plotted on the Y-axis, the relative protein amount is shown on the X-axis. As expected, all selected compounds exhibited a stabilization of CDK16 and GSK3 in the CETSA experiment. **21b**, **21e**, and **21h** revealed a broader range of stabilized kinase off-targets, which also correlated with the DSF data. Based on excellent proteome-wide selectivity identifying CDK16 and GSK3 as the most prominently stabilized kinases, **21d** and **21i** were selected for further investigations. In particular, **21i** showed excellent selectivity and potency in CETSA MS (Figure 6). Due to the high stabilization of CDK16 in the cellular system and the great selectivity, we hypothesised that we would see similar properties also for compounds **42d** and **43d** which had even more potent and selective properties from our biochemical screens. We therefore decided to proceed with a CDK16-specific cellular assay.
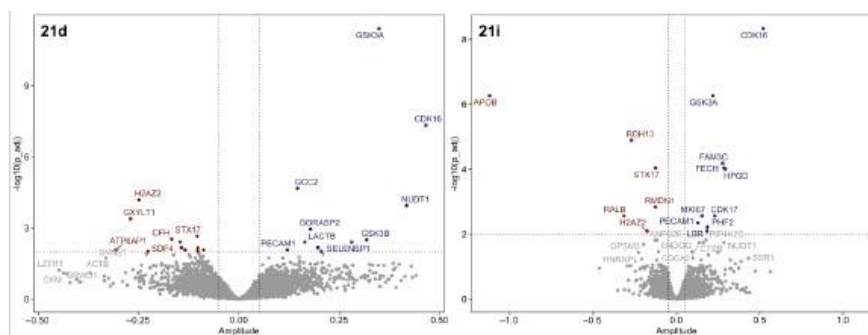
Figure 6. Volcano plots of stabilized (positive amplitude) and destabilized (negative amplitude) proteins of **21d** and **21i**.

2.4. NanoBRET Assay Revealed Highly Potent CDK Inhibitors in Cells

Next, to specifically quantify the outcome of the cellular activity of the compounds, a NanoBRET cellular target engagement assay was performed. The determination of the cellular potency on CDK16 correlated well with the corresponding DSF assay data (Figure S2). The lead structure (**1**) exhibited the highest DSF shift with 10.3 °C and a corresponding EC_{50} value of 18.0 nM (Figure 7). Compounds **11a–f** bearing the cyclopropyl moiety on the pyrazole ring resulted in excellent cellular activity with EC_{50} values of 33.0–124.0 nM. The introduction of an additional residue on the pyrimidine ring resulted in less potent inhibitors (**12a–e**, **13a**, **13c**), associated with a high number of potential off-targets. The second series yielded in some interesting compounds **21c**, **21e**, **21h**, and **21i** with EC_{50} values of 107.0, 152.0, 391.0, and 380.2 nM, respectively. Although slightly less potent than the lead structure and the derivatives from the first series with the cyclopropyl residue, these modifications resulted in more selective compounds with only 4–9 kinases identified with ΔTm shifts >5 °C (Figure 4). The introduction of a chlorine residue at the 5 position of the pyrimidine (**22a–22e**) and the exchange of the pyrimidine ring to a quinazoline (**23a–23c**) led to inactive compounds with EC_{50} values >1 μM on CDK16. As

mentioned before, the exchange from a methyl ester to an *iso*-propyl or *tert*-butyl ester led to a higher stabilization of the kinase in the DSF assay, which correlated with an increase in cellular potency as determined in the NanoBRET assay. **42d** and **43d** showed comparable DSF shifts to the lead structure and EC$_{50}$ values of 44.0 and 33.4 nM, respectively. The number of potential off-targets was low with 1 and 2 kinases with ΔTm >5 °C (Figure 4), highlighting these inhibitors for more comprehensive selectivity studied within the CDK and particular the TAIRE subfamily. The potency of **24** was also assessed for GSK3A and GSK3B in the NanoBRET assay, whereby a shift of 9.5 °C correlated with an EC$_{50}$ value of only 18 µM for GSK3B. **43d** exhibited a DSF shift of 2.9 °C and its activity against GSK3B was consequently negligible (Table S4).

Cmpd	R^1	R^2	R^4	NanoBRET EC$_{50}$ [nM] ± SEM		Cmpd	R^1	R^2	R^4	NanoBRET EC$_{50}$ [nM] ± SEM
1		H		18.0 ± 0.7		22a		Cl		1861.0 ± 276.0
11a		H		35.0 ± 0.0		22b		Cl		3188.0 ± 589.0
11b		H		124.0 ± 10.4		22c		Cl		8319.0 ± 227.5
11c		H		33.0 ± 0.1		22d		Cl		24980.0 ± 4220.0
11d		H		67.0 ± 1.3		22e		Cl		12085.0 ± 65.0
11e		H		36.0 ± 1.2		23a		Quin		9358.0 ± 1022.5
11f		H		110.0 ± 9.7		23b		Quin		6822.0 ± 279.5
12a		Me		172.0 ± 20.9		23c		Quin		4603.0 ± 385.0
12b		Me		1099.0 ± 44.5		24		H		2831.0 ± 544.5
12c		Me		176.0 ± 3.0		25		H		400.0 ± 30.6
12d		Me		619.0 ± 32.0		38a		H		388.0 ± 32.0
12e		Me		191.0 ± 5.9		38b		H		253.0 ± 19.2
13a		Cl		354.0 ± 60.5		39a		H		70.0 ± 4.9
13c		Cl		914.0 ± 11.0		39b		H		47.0 ± 4.1
15		H		100.0 ± 7.0		40a		H		1134.0 ± 20.5
21a		H		652.0 ± 15.6		40b		H		504.0 ± 9.3
21b		H		688.0 ± 22.5		40c		H		351.0 ± 34.1
21c		H		107.0 ± 3.0		41a		H		4165.0 ± 274.0
21d		H		1374.0 ± 306.0		41b		H		2467.0 ± 247.5
21e		H		152.0 ± 2.9		41c		H		5548.0 ± 636.0
21f		H		112.0 ± 8.6		42b		H		54.0 ± 1.8
21g		H		879.0 ± 5.5		42d		H		44.0 ± 7.5
21h		H		391.0 ± 46.4		43b		H		76.0 ± 0.9
21i		H		380.2 ± 230.2		43d		H		33.4 ± 5.8

Figure 7. Overview of the binding affinities of the *N*-(1*H*-pyrazole-3-yl)pyrimidine-4-amine-based inhibitors against CDK16. [a] EC$_{50}$ values were determined using the NanoBRET assay in a 11-point dose−response curve in duplicates.

To gain insights into CDK family-wide selectivity, we further profiled the most promising compounds **21i**, **42d**, and **43d**, as well as the lead structure **1** against the CDK family including co-expressed corresponding cyclins, using the NanoBRET technology. As expected, **1** was highly potent on all CDKs, but showed no selectivity within this kinase subfamily. **21i** exhibited an interesting profile, by targeting only CDK16 and CDK17 in the nanomolar range (380.2 and 673.1 nM). However, a much higher potency was observed for **42d** and **43d**, and **43d** exhibited the best selectivity profile within the CDK family. **43d** was highly potent for the PCTAIRE subfamily with 33.4 nM, 21.2 nM, and 120.6 nM for CDK16/cyclin Y, CDK17/cyclin Y, and CDK18/cyclin Y, respectively. The structurally highly related PFTAIRE subfamily was also targeted by **43d** with EC_{50} values ranging from 72.1 to 301.6 nM for CDK14/cyclin Y and CDK15/cyclin Y, respectively. Only CDK1, CDK9, and CDK10 were further off-targets inside the family but with weaker EC_{50} values of 581.9 nM, 993.4 nM, and 651.2 nM, respectively. **42d** showed additionally a nanomolar potency for CDK6, CDK7, and CDK20 (Figure 8, Table S5).
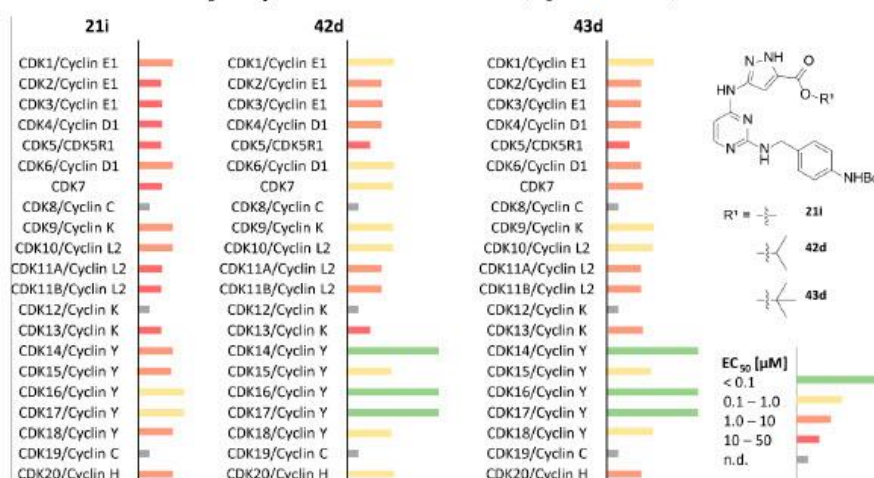
Figure 8. Cellular EC_{50} values of **21i**, **42d**, and **43d** against the CDK family were determined using NanoBRET technology in a 11-dose–response curve in duplicates.

2.5. Metabolic Stability of 21i and 43d

Since the most potent and selective PCTAIRE inhibitors harbored an ester linkage, we assessed metabolic stability using liver microsomes. The lead structure **1** was not included in this assay due to its promiscuous behaviour. Since **1** is presumably not suitable as a tool compound to study the effect of selective kinase inhibition in an in vivo application, we focused on the newly synthesized compounds **21i** and **43d**. The compounds were tested against an activated microsome mix derived from the liver of Sprague-Dawley rats, as previously reported [50]. **21i** and **43d** were incubated at 37 °C for 60 min and the amount of unmetabolized compound was determined in 15 min steps, using high-performance liquid chromatography (HPLC). These experiments revealed an increased metabolic stability for **43d** in comparison to **21i**, thus replacing the methyl ester in **21i** with a *tert*-butyl ester (Figure 9A). After an incubation time of 60 min, only 16% of **21i** were detected; however, microsomal stability was significantly increased to 79% for the *tert*-butyl ester **43d**.
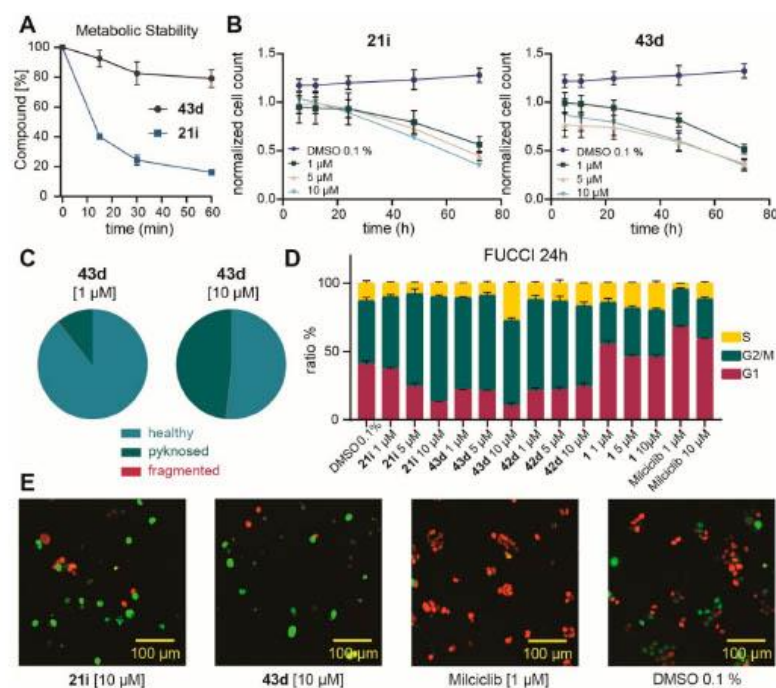
Figure 9. Metabolic stability and cell cycle analysis (FUCCI). (A). Metabolic stability of 21i [10 µM] and 43d [10 µM] after treatment of 60 min. The residual amount of compound in percent was plotted against time in minutes. (B). Normalized cell count of HCT116-FUCCI cells after 0, 6, 12, 24, 48 and 72 h of compound exposure (21i [1, 5, 10 µM], 43d [1, 5, 10 µM]) in comparison to cells exposed to 0.1 % DMSO. Error bars show SEM of two biological replicates. (C). Fraction of healthy, fragmented and pyknosed nuclei after 24 h of 1 and 10 µM compound exposure (43d) in HCT116 cells. Average data of two biological duplicates are shown. (D). Fractions of red (G1), green (G2/M) or yellow (S) cells after 24 h of compound exposure (21i [1 µM, 5 µM, 10 µM], 43d [1 µM, 5 µM, 10 µM], 42d [1, 5, 10 µM], 1 [1, 5, 10 µM], milciclib [1, 10 µM]) in comparison to cells exposed to 0.1 % DMSO. Error bars show SEM of two biological replicates. (E). Fluorescence Image of HCT116-FUCCI cells after 48 h of compounds exposure (21i [10 µM], 43d [10 µM]) in comparison to cells exposed to milciclib [1 µM] and cells exposed to 0.1 % DMSO.

2.6. Inhibition of CDK16 Led to Cell Cycle Arrest in G2/M

Expression of CDK16 has been shown to be cell cycle dependent with a peak in S and G2 phases [25]. Moreover, knockdown of CDK16 in different cancer cell lines like SCC has been shown to lead to late G2/M phase arrest followed by apoptosis, an effect that was mediated through regulation of the tumor suppressor p27 [28]. We therefore performed a cell-based assay in liver cells using the fluorescent ubiquitination-based cell cycle reporter (FUCCI) system [51,52]. This technology enables the detection of cell cycle states (G1, G2/M, or S phase) on a single cell level. Additionally, a viability assessment, as described by Tjaden et. al. [53], was performed using Hoechst33342 as a nuclear marker. Compounds 21i and 43d decreased the cell count at all concentrations tested (1, 5, 10 µM)

in a dose-dependent manner, in comparison to cells treated with DMSO (0.1 %) shortly after treatment (Figure 9B), and with an increasing effect over time. After 72 h, both compounds showed less than 50% cells compared to the control suggesting attenuated cell growth. The nuclear gating, based on a machine learning algorithm in healthy, pyknosed, and fragmented nuclei [53], revealed for cells treated with 1 μM of 43d on average 10% of pyknosed nuclei, an irreversible condensation of chromatin undergoing necrosis or apoptosis. At 10 μM of the same compound, the pyknosed nuclei rate was even higher, on average 45% (Figure 9C). An increase in cells including a pyknosed nucleus indicate apoptotic cell death, due to typical morphological changes like DNA condensation [54]. These results support the hypothesis of Yanagi et. al. that the inhibition of CDK16, which has been linked to up-regulation of p27, induces apoptosis. The compounds 21i, 42d, and 43d led to the accumulation of cells in G2/M phases at all three concentrations tested (1 μM, 5 μM, 10 μM) (Figure 9D,E). The most potent and selective inhibitor of our series, 43d, had the most significant impact on the cell cycle. Milciclib is a published CDK2 inhibitor and was used as a reference compound [55]. It led to G1 cell cycle arrest at both tested concentrations. In comparison, the lead structure compound 1, which displays no family or kinome wide-selectivity, had no effect on the cell cycle at all three concentrations tested. At later time points, the G2/M phase arrest was no longer detectable (Table S7). This can be explained due to the increasing amount of apoptotic cells. This observation is also consistent with CDK16 knockout data that have been shown to lead to a G2/M phase arrest followed by apoptotic cell death.

3. Discussion

We developed a new series of 3-amino-1*H*-pyrazole-based kinase inhibitors, based on the scaffold of the promiscuous inhibitor 1. Selectivity screens against a representative set of 100 kinases revealed that small modifications, especially on the pyrazole ring, had significant effects on the selectivity of the new synthesized compounds. Alkyl residues on the pyrazole led to non-selective inhibitors, whereby the introduction of an amide moiety instead of the ester linkage was not tolerated and yielded in inactive compounds. The introduction of a methyl ester caused a dramatic reduction of potential hits while maintaining high potency on CDK16. The replacement of the methyl ester to an *iso*-propyl or *tert*-butyl ester increased the stabilization on CDK16 and led to a great selectivity profile in our in-house DSF panel, screening over one hundred kinases representing the human kinome. The cellular potency against CDK16 was determined for all new synthesized compounds, using the NanoBRET technology. The optimization of the promiscuous scaffold led to 42d and 43d, which were highly potent on CDK16 with a comparable potency to staurosporine or the lead structure 1, which had reported low nM activity in enzyme-based assays. Further profiling against the whole CDK family revealed that 43d only targeted the PCTAIRE and PFTAIRE subfamily with low nanomolar activity in cells. All members of this family belong to the dark kinome of understudied kinases. 43d showed good metabolic stability at the compound induced cell cycle arrest and apoptosis in agreement with genetic knockout studies of CDK16. Thus, the reported synthesis and discovery of the tool compound 43d provides a well characterized inhibitor for cell-based mechanistic studies on the PCTAIRE family of CDKs.

4. Materials and Methods

4.1. Differential Scanning Fluorimetry Assay

Recombinant protein kinase domains with a concentration of 2 μM were mixed with a 10 μM compound solution in DMSO, 20 mM HEPES, pH 7.5, and 500 mM NaCl. SYPRO Orange (5000×, Invitrogen) was added as a fluorescence probe (1 μl per mL). Subsequently, temperature-dependent protein unfolding profiles were measured, using the QuantStudio™ 5 realtime PCR machine (Thermo Fisher, Waltham, MA, USA). Excitation and emission filters were set to 465 and 590 nm. The temperature was raised with a step rate of 3 °C per minute. Data points were analyzed with the internal software (Thermal Shift Software™

Version 1.4, Thermo Fisher) using the Boltzmann equation to determine the inflection point of the transition curve [56]. Differences in melting temperature are given as ΔTm values in °C. Measurements were performed in duplicates.

4.2. Cellular Thermal Shift Assay

Compressed CETSA MS assays, also known as PISA (Proteome Integral Stability Alteration assay), were performed as previously described by Chernobrovkin et. al., with some minor changes [57]. K562 cells were obtained from ATCC. K562 cells were mixed with an equal volume of compound solution, final concentartion of 30 µM, in a buffer (20 mM HEPES, 138 mM NaCl, 5 mM KCl, 2 mM CaCl$_2$, 1 mM MgCl$_2$, pH 7.4). One percent and DMSO was used as a control. Cells were incubated at +37 °C for 60 min. The cell suspension was split into 12 aliquots which were heated at a different temperature between 44 and 66 °C for 3 min. After the heating, the cells were lysed through freeze–thawing them 3 times and the lysed aliquots were pooled together. Cell debris and aggregates were removed through centrifugation (20 min at 30.000× g).

4.3. NanoBRET

The assay was performed as described previously [58]. In brief: Full-length kinases were obtained as plasmids cloned in frame with a terminal NanoLuc-fusion (Promega, Madison, WI, USA), as specified in table (Table S6). Plasmids were transfected into HEK293T cells using FuGENE HD (Promega, E2312) and proteins were allowed to express for 20h. Serially diluted inhibitor and NanoBRET Kinase Tracer (Promega) at a concentration determined previously as the $K_{D,app}$ (Table S5) were pipetted into white 384-well plates (Greiner 781207) using an Echo acoustic dispenser (Labcyte, San Jose, CA, USA). The corresponding protein-transfected cells were added and reseeded at a density of 2.5×10^5 cells/mL after trypsinization and resuspending in Opti-MEM without phenol red (Life Technologies, Carlsbad, CA, USA). The system was allowed to equilibrate for 2 h at 37 °C/5% CO$_2$ prior to BRET measurements. To measure BRET, NanoBRET NanoGlo Substrate + Extracellular NanoLuc Inhibitor (Promega, N2540) was added as per the manufacturer's protocol, and filtered luminescence was measured on a PHERAstar plate reader (BMG Labtech, Ortenberg, Germany) equipped with a luminescence filter pair (450 nm BP filter (donor) and 610 nm LP filter (acceptor)). Competitive displacement data were then graphed using GraphPad Prism 9 software using a normalized 3-parameter curve fit with the following equation: Y = 100/(1 + 10^(X−LogEC50)).

4.4. Microsomal Stability Assay

The solubilized test compound (5 µL, final concentration 10 µM) was preincubated at 37 °C in 432 µL of phosphate buffer (0.1 M, pH 7.4) together with 50 µL NADPH regenerating system (30 mM glucose 6 phosphate, 4 U/mL glucose 6 phosphate dehydrogenase, 10 mM NADP, 30 mM MgCl$_2$). After 5 min, the reaction was started by the addition of 13 µL of microsome mix from the liver of Sprague−Dawley rats (Invitrogen; 20 mg protein/mL in 0.1 M phosphate buffer) in a shaking water bath at 37 °C. The reaction was stopped by adding 500 µL of ice cold methanol at 0, 15, 30 and 60 min. The samples were centrifuged at 5000× g for 5 min at 4 °C, and the supernatants were analyzed and test compound was quantified by HPLC: The composition of the mobile phase is adapted to the test compound in a range of MeOH 40–90% and water (0.1% formic acid) 10–60%; flow rate: 1 mL/min; stationary phase: Purospher® STAR, RP18, 5 µm, 125 × 4, precolumn: Purospher® STAR, RP18, 5 µm, 4 × 4; detection wavelength: 254 and 280 nm; injection volume: 50 µL. Control samples were performed to check the test compound's stability in the reaction mixture: first control was without NADPH, which is needed for the enzymatic activity of the microsomes, second control was with inactivated microsomes (incubated for 20 min at 90 °C), and third control was without test compound (to determine the baseline). The amounts of the test compound were quantified by an external calibration curve. Data are expressed as the mean ± SEM remaining compound from three independent experiments.

4.5. FUCCI Cell Cycle Assay

To validate the influence of the compounds on the cell cycle, a florescent ubiquitination-based cell cycle indicator (FUCCI) assay was performed as described previously [53]. In Brief, HCT116-FUCCI cells, stably expressing the FUCCI system, introduced by the Sleeping Beauty transposon system [53,59], were seeded at a density of 1250 cells per well in a 384 well plate (Cell culture microplate, PS, f-bottom, µClear®, 781091, Greiner, Frickenhausen, Germany) in culture medium (50 µL per well) and stained additionally with 60 nM Hoechst33342 (Thermo Scientific). Fluorescence and cellular shape were measured before and after compound treatment for 72 h every 12 h using the CQ1 high-content confocal microscope (Yokogawa, Tokyo, Japan). Compounds were added directly to the cells at three different concentrations (1 µM, 5 µM and 10 µM). The following parameters were used for image acquisition: Ex 405 nm/Em 447/60 nm, 500ms, 50%; Ex 561 nm/Em 617/73 nm, 100 ms, 40%; Ex 488/Em 525/50 nm, 50 ms, 40%; Ex 640 nm/Em 685/40, 50 ms, 20 %; bright field, 300ms, 100% transmission, one centered field per well, 7 z stacks per well with 55 µm spacing. Analysis of images was performed using the CellPathfinder software (Yokogawa). The cell count was normalized against the cell count of cells treated with 0.1% DMSO. All normal gated cells were further classified in cells containing healthy, fragmented or pyknosed nuclei. Cells that showed a healthy nucleus were gated in red, green or yellow based on 11 features of the cellbody and 4 features of the cell nuclei. Error bars show SEM of biological duplicates. All data can be found in Table S7.

4.6. Chemistry

The synthesis of compounds will be explained in the following and the analytical data for them can be found in the Supporting Information. All commercial chemicals were purchased from common suppliers with a purity ≥95% and were used without further purification. The solvents with an analytical grade were obtained from VWR Chemicals and Merck and all dry solvents from Acros Organics. All reactions were proceeded under an argon atmosphere. The thin layer chromatography was done with silica gel on aluminium foils (60 Å pore diameter) obtained from Macherey-Nagel and visualized with ultraviolet light (λ = 254 and 365 nm). The purification of the compounds was done by flash chromatography. A puriFlash XS 420 device with a UV-VIS multiwave detector (200−400 nm) from Interchim was used with pre-packed normal-phase PF-SIHP silica columns with particle sizes of 15 and 30 µm (Interchim). The nuclear magnetic resonance spectroscopy (NMR) was performed on a DPX250, AV300, AV400 or AV500 MHz spectrometers from Bruker. Chemical shifts (δ) are reported in parts per million (ppm). DMSO-d6 was used as a solvent, and the spectra were calibrated to the solvent signal: 2.50 ppm (^{1}H NMR) or 39.52 ppm (^{13}C NMR) for DMSO-d6. Coupling constants (J) were reported in hertz (Hz) and multiplicities were designated as followed: s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet). Mass spectra were measured on a Surveyor MSQ device from ThermoFisher measuring in the positive- or negative-ion mode. Final compounds were additionally characterized by HRMS using a MALDI LTQ Orbitrap XL from ThermoScientific. The purity of the final compounds was determined by HPLC using method A: an Agilent 1260 Infinity II device with a 1260 DAD HS detector (G7117C; 254 nm, 280 nm, 310 nm) and a LC/MSD device (G6125B, ESI pos. 100–1000). The compounds were analyzed on a Poroshell 120 EC-C18 (Agilent, 3 × 150 mm, 2.7 µm) reversed phase column using 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B) as a mobile phase. The following gradient was used: 0 min 5% B–2 min 5% B–8 min 98% B–10 min 98% B (flow rate of 0.5 mL/min). Method B: an Agilent 1260 Infinity II device, with a 1260 MWD detector (G7165A; 254 nm, 280 nm) and a LC/MSD device (G6125B, ESI pos. 100–1000) was used. The compounds were analyzed on an Eclipse XDB-C18 (Agilent, 4.6 × 250 mm, 5 µm) reversed phase column using 0.1% TFA in water (A) and 0.1% TFA in acetonitrile (B) as a mobile phase. The following gradient was used: 0 min. 2% B–2 min. 2% B–10 min. 98% B–15 min. 98% B–17 min. 2% B–19 min. 2% B (flow rate of 1 mL/min.). UV-detection

was performed at 254, 280 and 310 nm and all compounds used for further biological characterizations showed a purity ≥95%.

General procedure 1. The 3-aminopyrazole derivative (1.1 equiv) and the corresponding 2,4-dichloropyrimidine (1.0 equiv) were dissolved in anhydrous isopropanol (0.15 M). TEA (3.0 equiv) was added and the mixture was stirred at 50–80 °C for 18–120 h. The solvent was evaporated under reduced pressure and the crude product was purified by flash chromatography using DCM/methanol or n-hexane/EE as an eluent.

General procedure 2. The corresponding product of general procedure 1 (1.0 equiv) and the aniline derivative (1.0 equiv) were dissolved in anhydrous ethanol (0.07 M). A catalytic amount of 1 M HCl was added and the mixture was stirred at 70 °C–reflux for 4–18 h. A solid precipitated, which was filtered and washed with ethanol to obtain the title compound.

General procedure 3. The corresponding product of general procedure 1 (1.0 equiv) and the aniline derivative (1.0 equiv) were dissolved in anhydrous ethanol (0.07 M). A catalytic amount of 1 M HCl was added and the mixture was stirred at 70 °C–reflux for 18–24 h. The solvent was evaporated under reduced pressure and the crude product was purified by flash chromatography using DCM/methanol or n-hexane/EE as an eluent.

General procedure 4. The corresponding product of general procedure 1 (1.0 equiv) and the primary amine (1.0 equiv) were dissolved in anhydrous ethanol (0.07 M). TEA (3.0 equiv) was added and the mixture was stirred at 80–120 °C for 3–10 h under microwave irradiation. The solvent was evaporated under reduced pressure and the crude product was purified by flash chromatography using DCM/methanol or n-hexane/EE as an eluent.

2-bromo-N-(3-cyclopropyl-1H-pyrazol-5-yl)pyrimidin-4-amine (**8**). The title compound was prepared according to the general procedure 1, using 5-cyclopropyl-1H-pyrazol-3-amine and 2,4-dichloropyrimidine. The mixture was stirred at 55 °C for 48 h to obtain the product (133 mg, 35%) as a white solid. 1H NMR (250 MHz, DMSO-d6) δ 12.17 (s, 1H), 10.26 (s, 1H), 8.08 (d, J = 5.9 Hz, 1H), 7.26 (s, 1H), 5.96 (s, 1H), 1.95–1.81 (m, 1H), 0.98–0.88 (m, 2H), 0.73–0.62 (m, 2H). 13C NMR (126 MHz, DMSO) δ 160.36, 158.81, 151.52, 151.40, 147.27, 105.57, 92.64, 7.75, 6.69. MS-ESI m/z [M + H]$^+$: calcd 281.1, found 280.0.

2-chloro-N-(3-cyclopropyl-1H-pyrazol-5-yl)-5-methylpyrimidin-4-amine (**9**). The title compound was prepared according to the general procedure 1, using 5-cyclopropyl-1H-pyrazol-3-amine and 2,4-dichloro-5-methylpyrimidine. The mixture was stirred at 80 °C for 48 h to obtain the product (366 mg, 40%) as a beige solid. ^{1}H NMR (300 MHz, DMSO-d6) δ 12.17 (s, 1H), 9.27 (s, 1H), 7.97 (d, J = 1.0 Hz, 1H), 6.26 (s, 1H), 2.11 (d, J = 1.0 Hz, 3H), 1.97–1.85 (m, 1H), 0.98–0.89 (m, 2H), 0.74–0.66 (m, 2H). ^{13}C NMR (75 MHz, DMSO-d6) δ 159.78, 156.99, 155.86, 113.90, 94.84, 13.37, 7.66, 6.97. MS-ESI m/z [M–H]$^-$: calcd 248.7, found 248.0.

2,5-dichloro-N-(3-cyclopropyl-1H-pyrazol-5-yl)pyrimidin-4-amine (**10**). The title compound was prepared according to the general procedure 1, using 5-cyclopropyl-1H-pyrazol-3-amine and 2,4,5-trichloropyrimidine. The mixture was stirred at 60 °C for 48 h to obtain the product (866 mg, 89%) as a white solid. 1H NMR (300 MHz, DMSO-d6) δ 12.33 (s, 1H), 9.65 (s, 1H), 8.32 (s, 1H), 6.19 (s, 1H), 1.99–1.82 (m, 1H), 0.99–0.89 (m, 2H), 0.74–0.66 (m, 2H). 13C NMR (75 MHz, DMSO) δ 157.05, 156.86, 155.14, 145.81, 145.55, 113.13, 95.57, 7.72, 6.79. MS-ESI m/z [M–H]$^-$: calcd 269.1, found 268.1.

$tert$-butyl 4-((4-((3-cyclopropyl-1H-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)phenethyl-carbamate (**11a**). The title compound was prepared according to the general procedure 2, using 5 and $tert$-butyl 4-aminophenethylcarbamate. The mixture was stirred for 4 h under reflux to obtain the product (155 mg, 67%) as a white solid. 1H NMR (300 MHz, DMSO-d6) δ 12.42 (s, 1H), 11.18 (s, 1H), 10.39 (s, 1H), 7.91 (d, J = 5.8 Hz, 1H), 7.40 (d, J = 7.8 Hz, 2H), 7.27 (d, J = 8.2 Hz, 2H), 6.91 (t, J = 5.4 Hz, 1H), 6.43 (s, 1H), 6.06 (s, 1H), 3.17 (q, J = 6.9 Hz, 2H), 2.74 (t, J = 7.5 Hz, 2H), 1.90–1.74 (m, 1H), 1.37 (s, 9H), 0.99–0.87 (m, 2H), 0.54 (s, 2H). 13C NMR (75 MHz, DMSO) δ 159.54, 155.52, 152.68, 152.56, 149.03, 145.91, 137.11, 134.25, 129.20, 123.96, 99.20, 93.29, 77.53, 41.44, 35.06, 28.24, 7.88, 6.70. MS-ESI m/z [M + H]$^+$: calcd

436.5, found 436.4. HRMS m/z [M + H]$^+$: calcd 436.2456, found 436.2446. HPLC: t_R = 7.57, purity $\geq$ 95% (UV: 254/280 nm).

tert-butyl 3-((4-((3-cyclopropyl-1*H*-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)phenethyl-carbamate (**11b**). The title compound was prepared according to the general procedure 3, using **5** and *tert*-butyl 3-aminophenethylcarbamate. The mixture was stirred for 18 h under reflux to obtain the product (29 mg, 19%) as a colorless oil. 1H NMR (300 MHz, DMSO-d6) δ 11.97 (s, 1H), 9.48 (s, 1H), 8.97 (s, 1H), 7.98 (d, J = 5.8 Hz, 1H), 7.66 (d, J = 8.2 Hz, 1H), 7.50 (s, 1H), 7.16 (t, J = 7.8 Hz, 1H), 6.86 (t, J = 5.6 Hz, 1H), 6.75 (d, J = 7.5 Hz, 1H), 6.58–6.09 (m, 2H), 3.21–3.07 (m, 2H), 2.66 (t, J = 7.6 Hz, 2H), 1.93–1.78 (m, 1H), 1.37 (s, 9H), 0.89 (d, J = 8.2 Hz, 2H), 0.74–0.62 (m, 2H). 13C NMR (75 MHz, DMSO) δ 159.61, 155.90, 155.54, 140.98, 139.54, 128.28, 121.20, 119.14, 116.81, 98.14, 92.99, 77.51, 48.61, 41.66, 35.89, 28.27, 7.66. MS-ESI m/z [M + H]$^+$: calcd 436.5, found 436.4. HRMS m/z [M + H]$^+$: calcd 436.2456, found 436.2452. HPLC: t_R = 7.58, purity $\geq$ 95% (UV: 254/280 nm).

tert-butyl (5-((4-((3-cyclopropyl-1*H*-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)pentyl) carbamate (**11c**). The title compound was prepared according to the general procedure 4, using **5** and *tert*-butyl (5-aminopentyl)carbamate. The mixture was stirred for 5 h at 90 °C to obtain the product (37 mg, 42%) as a colorless oil. 1H NMR (300 MHz, DMSO-d6) δ 11.92 (s, 1H), 9.31 (s, 1H), 7.77 (d, J = 5.7 Hz, 1H), 6.74 (t, J = 5.9 Hz, 1H), 6.65 (s, 1H), 6.30–5.97 (m, 2H), 3.21 (q, J = 6.7 Hz, 2H), 2.98–2.84 (m, 2H), 1.91–1.76 (m, 1H), 1.61–1.43 (m, 2H), 1.43–1.33 (m, 11H), 1.33–1.20 (m, 2H), 0.99–0.82 (m, 2H), 0.76–0.60 (m, 2H). 13C NMR (75 MHz, DMSO) δ 162.30, 162.02, 155.99, 155.57, 123.51, 95.71, 77.29, 40.66, 35.77, 30.76, 29.33, 28.98, 28.26, 23.93, 7.65. MS-ESI m/z [M + H]$^+$: calcd 402.4, found 402.2. HRMS m/z [M + H]$^+$: calcd 402.2612, found 402.2622. HPLC: t_R = 7.37, purity $\geq$ 95% (UV: 254/280 nm).

tert-butyl (2-(2-((4-((3-cyclopropyl-1*H*-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)ethoxy)-ethyl)carbamate (**11d**). The title compound was prepared according to the general procedure 4, using **5** and *tert*-butyl (2-(2-aminoethoxy)ethyl)carbamate. The mixture was stirred for 7 h at 90 °C to obtain the product (43 mg, 30%) as a colorless oil. 1H NMR (300 MHz, DMSO-d6) δ 11.92 (s, 1H), 9.34 (s, 1H), 7.79 (d, J = 5.7 Hz, 1H), 6.92–6.50 (m, 2H), 6.30–6.00 (m, 2H), 3.56–3.46 (m, 2H), 3.45–3.36 (m, 4H), 3.09 (q, J = 5.9 Hz, 2H), 1.91–1.76 (m, 1H), 1.37 (s, 9H), 0.95–0.83 (m, 2H), 0.72–0.63 (m, 2H). 13C NMR (75 MHz, DMSO) δ 161.92, 159.58, 156.01, 155.61, 95.89, 92.13, 77.61, 69.12, 40.47, 28.23, 7.84, 7.68. MS-ESI m/z [M + H]$^+$: calcd 404.5, found 404.4. HRMS m/z [M + H]$^+$: calcd 404.2405, found 404.2405. HPLC: t_R = 5.46, purity $\geq$ 95% (UV: 254/280 nm).

tert-butyl (6-((4-((3-cyclopropyl-1*H*-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)hexyl)-carbamate (**11e**). The title compound was prepared according to the general procedure 4, using **5** and *tert*-butyl (6-aminohexyl)carbamate. The mixture was stirred for 7 h at 90 °C to obtain the product (66 mg, 45%) as a colorless oil. 1H NMR (300 MHz, DMSO-d6) δ 11.87 (s, 1H), 9.29 (s, 1H), 7.77 (d, J = 5.6 Hz, 1H), 6.73 (t, J = 5.7 Hz, 1H), 6.62 (s, 1H), 6.35–6.00 (m, 2H), 3.44–3.28 (m, 2H), 3.28–3.14 (m, 2H), 2.89 (q, J = 6.5 Hz, 2H), 1.89–1.77 (m, 1H), 1.56–1.44 (m, 2H), 1.36 (s, 9H), 1.33–1.22 (m, 4H), 0.97–0.83 (m, 2H), 0.73–0.60 (m, 2H). 13C NMR (75 MHz, DMSO) δ 162.06, 159.61, 156.03, 155.59, 95.62, 77.28, 69.07, 41.18, 40.67, 29.57, 29.33, 28.26, 26.39, 26.20, 7.66. MS-ESI m/z [M + H]$^+$: calcd 416.5, found 416.5. HRMS m/z [M + H]$^+$: calcd 416.2769, found 416.2768. HPLC: t_R = 7.557, purity $\geq$ 95% (UV: 254/280 nm).

tert-butyl 4-(((4-((3-cyclopropyl-1*H*-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)methyl)-benzylcarbamate (**11f**). The title compound was prepared according to the general procedure 4, using **5** and *tert*-butyl 4-(aminomethyl)benzylcarbamate. The mixture was stirred for 6 h at 90 °C to obtain the product (22 mg, 24%) as a colorless oil. 1H NMR (300 MHz, DMSO-d6) δ 11.97 (s, 1H), 9.35 (s, 1H), 7.79 (d, J = 5.7 Hz, 1H), 7.38–7.10 (m, 6H), 6.26–5.92 (m, 2H), 4.46 (d, J = 6.2 Hz, 2H), 4.08 (d, J = 6.2 Hz, 2H), 1.88–1.73 (m, 1H), 1.38 (s, 9H), 0.93–0.79 (m, 2H), 0.71–0.53 (m, 2H). 13C NMR (75 MHz, DMSO) δ 161.99, 159.65, 156.01, 155.77, 139.25, 138.28, 126.81, 96.08, 77.69, 43.72, 43.15, 28.24, 7.65. MS-ESI m/z [M + H]$^+$: calcd 436.5, found 436.4. HRMS m/z [M + H]$^+$: calcd 436.2456, found 436.2448. HPLC: t_R = 7.45, purity $\geq$ 95% (UV: 254/280 nm).

tert-butyl 4-((4-((3-cyclopropyl-1*H*-pyrazol-5-yl)amino)-5-methylpyrimidin-2-yl)amino)-phenethylcarbamate (**12a**). The title compound was prepared according to the general procedure 3, using **6** and *tert*-butyl 4-aminophenethylcarbamate. The mixture was stirred for 18 h under reflux to obtain the product (85 mg, 90%) as a colorless oil. 1H NMR (300 MHz, DMSO-d6) δ 12.56 (s, 1H), 10.61 (s, 1H), 10.16 (s, 1H), 7.88 (s, 1H), 7.38 (d, J = 7.2 Hz, 2H), 7.19 (d, J = 7.3 Hz, 2H), 6.90 (s, 1H), 6.07 (s, 1H), 3.21–3.07 (m, 2H), 2.77–2.63 (m, 2H), 2.11 (s, 3H), 1.94–1.81 (m, 1H), 0.99–0.87 (m, 2H), 0.64–0.51 (m, 2H). 13C NMR (75 MHz, DMSO) δ 160.12, 155.41, 151.27, 146.40, 145.12, 140.12, 136.01, 134.68, 128.98, 122.29, 107.06, 95.10, 77.40, 41.38, 34.96, 28.16, 13.26, 7.84, 6.78. MS-ESI *m/z* [M + H]⁺: calcd 450.6, found 450.4. HRMS *m/z* [M + H]⁺: calcd 450.2612, found 450.2603. HPLC: t$_R$ = 7.54, purity ≥ 95% (UV: 254/280 nm).

tert-butyl 3-((4-((3-cyclopropyl-1*H*-pyrazol-5-yl)amino)-5-methylpyrimidin-2-yl)amino)-phenethylcarbamate (**12b**). The title compound was prepared according to the general procedure 3, using **6** and *tert*-butyl 3-aminophenethylcarbamate. The mixture was stirred for 18 h under reflux to obtain the product (34 mg, 36%) as a colourless oil. 1H NMR (300 MHz, DMSO-d6) δ 12.09 (s, 1H), 9.12 (s, 2H), 7.86 (s, 1H), 7.62 (d, J = 8.2 Hz, 1H), 7.45 (s, 1H), 7.14 (t, J = 7.8 Hz, 1H), 6.86 (t, J = 5.7 Hz, 1H), 6.72 (d, J = 7.5 Hz, 1H), 5.83 (s, 1H), 3.33 (s, 3H), 3.21–3.05 (m, 2H), 2.72–2.59 (m, 2H), 1.93–1.79 (m, 1H), 1.37 (s, 9H), 0.96–0.81 (m, 2H), 0.72–0.61 (m, 2H). 13C NMR (75 MHz, DMSO) δ 158.12, 155.52, 155.14, 141.21, 139.56, 128.35, 120.81, 118.44, 116.14, 105.46, 77.49, 41.65, 35.91, 28.26, 13.33, 7.68. MS-ESI *m/z* [M + H]⁺: calcd 450.6, found 450.4. HRMS *m/z* [M + H]⁺: calcd 450.2612, found 450.2607. HPLC: t$_R$ = 7.63, purity ≥ 95% (UV: 254/280 nm).

tert-butyl (5-((4-((3-cyclopropyl-1*H*-pyrazol-5-yl)amino)-5-methylpyrimidin-2-yl)amino)-pentyl)carbamate (**12c**). The title compound was prepared according to the general procedure 4, using **6** and *tert*-butyl (5-aminopentyl)carbamate. The mixture was stirred for 6 h at 90 °C to obtain the product (5 mg, 5%) as a colorless oil. 1H NMR (300 MHz, DMSO-d6) δ 12.06 (s, 1H), 8.58 (s, 1H), 7.65 (s, 1H), 6.75 (t, J = 5.5 Hz, 1H), 6.32 (s, 1H), 3.18 (q, J = 6.7 Hz, 2H), 2.90 (q, J = 6.5 Hz, 2H), 1.96 (s, 3H), 1.90–1.78 (m, 1H), 1.55–1.45 (m, 2H), 1.36 (s, 9H), 1.32–1.21 (m, 4H), 0.97–0.81 (m, 2H), 0.69–0.61 (m, 2H). MS-ESI *m/z* [M + H]⁺: calcd 416.5, found 416.2. HRMS *m/z* [M + H]⁺: calcd 416.2769, found 416.2769. HPLC: t$_R$ = 7.43, purity ≥ 95% (UV: 254/280 nm).

tert-butyl (2-(2-(4-((3-cyclopropyl-1*H*-pyrazol-5-yl)amino)-5-methylpyrimidin-2-yl)amino)ethoxy)ethyl)carbamate (**12d**). The title compound was prepared according to the general procedure 4, using **6** and *tert*-butyl (2-(2-aminoethoxy)ethyl)carbamate. The mixture was stirred for 10 h at 90 °C to obtain the product (7 mg, 7%) as a colorless oil. 1H NMR (300 MHz, DMSO-d6) δ 8.66 (s, 1H), 7.66 (s, 1H), 6.92–6.70 (m, 2H), 6.15 (s, 1H), 3.49 (t, J = 5.8 Hz, 2H), 3.44–3.37 (m, 2H), 3.08 (q, J = 5.9 Hz, 2H), 2.94–2.84 (m, 2H), 1.97 (s, 3H), 1.90–1.78 (m, 1H), 1.37 (s, 9H), 0.96–0.79 (m, 2H), 0.71–0.62 (m, 2H). 13C NMR (75 MHz, DMSO) δ 160.77, 155.55, 116.98, 102.86, 77.61, 69.28, 69.12, 41.33, 40.57, 28.22, 23.63, 13.19, 7.70. MS-ESI *m/z* [M + H]⁺: calcd 418.5, found 418.4. HRMS *m/z* [M + H]⁺: calcd 418.2561, found 418.2559. HPLC: t$_R$ = 7.18, purity ≥ 95% (UV: 254/280 nm).

tert-butyl (6-((4-((3-cyclopropyl-1*H*-pyrazol-5-yl)amino)-5-methylpyrimidin-2-yl)amino)hexyl)carbamate (**12e**). The title compound was prepared according to the general procedure 4, using **6** and *tert*-butyl (6-aminohexyl)carbamate. The mixture was stirred for 6 h at 90 °C to obtain the product (5 mg, 5%) as a colourless oil. 1H NMR (300 MHz, DMSO-d6) δ 12.17 (s, 1H), 8.57 (s, 1H), 7.67 (s, 1H), 6.78 (t, J = 5.0 Hz, 1H), 6.44 (s, 1H), 3.21 (q, J = 6.7 Hz, 2H), 2.91 (q, J = 6.5 Hz, 2H), 1.98 (s, 3H), 1.92–1.80 (m, 1H), 1.60–1.47 (m, 2H), 1.38 (s, 9H), 1.33–1.24 (m, 4H), 0.96–0.82 (m, 2H), 0.71–0.60 (m, 2H). MS-ESI *m/z* [M + H]⁺: calcd 430.6, found 430.3. HRMS *m/z* [M + H]⁺: calcd 430.2925, found 430.2924. HPLC: t$_R$ = 7.69, purity ≥ 95% (UV: 254/280 nm).

tert-butyl 4-((5-chloro-4-((3-cyclopropyl-1*H*-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)-phenethylcarbamate (**13a**). The title compound was prepared according to the general procedure 3, using **7** and *tert*-butyl 4-aminophenethylcarbamate. The mixture was stirred for 18 h under reflux to obtain the product (20 mg, 16%) as a yellow oil. 1H NMR (300 MHz,

DMSO-d6) δ 12.25 (s, 1H), 9.31 (s, 1H), 8.58 (s, 1H), 8.09 (s, 1H), 7.54 (d, J = 8.4 Hz, 2H), 7.07 (d, J = 8.1 Hz, 2H), 6.85 (t, J = 5.6 Hz, 1H), 6.16 (s, 1H), 3.11 (q, J = 6.9 Hz, 2H), 2.63 (t, J = 7.6 Hz, 2H), 2.00–1.80 (m, 1H), 1.37 (s, 9H), 0.98–0.86 (m, 2H), 0.74–0.62 (m, 2H). 13C NMR (75 MHz, DMSO) δ 157.86, 155.52, 155.10, 154.26, 138.41, 132.39, 128.52, 119.32, 103.36, 77.46, 41.73, 34.99, 28.26, 7.73. MS-ESI *m/z* [M + H]⁺: calcd 471.0, found 470.4. HRMS *m/z* [M + H]⁺: calcd 470.2066, found 470.2061. HPLC: t_R = 8.70, purity ≥ 95% (UV: 254/280 nm).

tert-butyl (5-((5-chloro-4-((3-cyclopropyl-1*H*-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)pentyl)carbamate (**13c**). The title compound was prepared according to the general procedure 4, using 7 and *tert*-butyl (5-aminopentyl)carbamate. The mixture was stirred for 8 h at 80 °C to obtain the product (50 mg, 52%) as a colorless oil. 1H NMR (300 MHz, DMSO-d6) δ 12.11 (s, 1H), 8.42 (s, 1H), 7.91 (s, 1H), 7.02 (s, 1H), 6.74 (t, J = 5.6 Hz, 1H), 6.31 (s, 1H), 3.27–3.13 (m, 2H), 2.90 (q, J = 6.5 Hz, 2H), 1.93–1.80 (m, 1H), 1.58–1.46 (m, 2H), 1.43–1.34 (m, 11H), 1.33–1.21 (m, 2H), 0.98–0.85 (m, 2H), 0.73–0.61 (m, 2H). 13C NMR (75 MHz, DMSO) δ 160.41, 158.17, 155.59, 154.56, 101.48, 93.37, 77.29, 40.99, 29.30, 28.77, 28.26, 23.87, 7.69. MS-ESI *m/z* [M + H]⁺: calcd 437.0, found 436.4. HRMS *m/z* [M + H]⁺: calcd 436.2222, found 436.2218. HPLC: t_R = 7.56, purity ≥ 95% (UV: 254/280 nm).

N^2-(6-aminohexyl)-N^4-(3-cyclopropyl-1*H*-pyrazol-5-yl)pyrimidine-2,4-diamine (**14**). 8e (50 mg, 0.1 mmol) was dissolved in anhydrous DCM (4 mL). TFA (549 mg, 4.8 mmol) was added at 0°C and the reaction mixture was allowed to warm up to rt overnight. The solvent was evaporated under reduced pressure. The residue was dissolved in methanol and neutralized with saturated K_2CO_3 solution. The solvent was again evaporated under reduced pressure and the crude product was purified by flash chromatography using H_2O/acetonitrile as an eluent to obtain the desired product with impurities. The product was used without further purification. MS-ESI *m/z* [M + H]⁺: calcd 316.4, found 316.2.

N-(6-((4-((3-cyclopropyl-1*H*-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)hexyl)acetamide (**15**). Acetic acid (7 μL, 0.1 mmol) and HATU (55 mg, 0.1 mmol) were dissolved in anhydrous DMF (4 mL). DIPEA (37 mg 0.3 mmol) was added and the resulting mixture was stirred at rt for 1 h. **11** (38 mg, 0.1 mmol) was added and the reaction mixture was stirred at rt for further 18 h. The solvent was evaporated under reduced pressure and the crude product was purified by flash chromatography using DCM/ethanol as an eluent to obtain the desired product (13 mg, 30%) as a colorless oil. 1H NMR (300 MHz, DMSO-d6) δ 12.42 (s, 1H), 11.12 (s, 1H), 8.47 (s, 1H), 7.92–7.68 (m, 2H), 6.57–6.17 (m, 2H), 3.39–3.21 (m, 2H), 3.00 (q, J = 6.4 Hz, 2H), 1.97–1.82 (m, 1H), 1.77 (s, 3H), 1.67–1.50 (m, 2H), 1.47–1.23 (m, 6H), 1.02–0.90 (m, 2H), 0.83–0.62 (m, 2H). 13C NMR (75 MHz, DMSO) δ 168.91, 153.77, 150.30, 146.28, 142.21, 97.72, 93.29, 40.93, 38.41, 29.14, 28.36, 26.17, 22.60, 7.86, 6.76. MS-ESI *m/z* [M + H]⁺: calcd 358.5, found 358.3. HRMS *m/z* [M + H]⁺: calcd 358.2350, found 358.2352. HPLC: t_R = 6.51, purity ≥ 95% (UV: 254/280 nm).

methyl 5-((2-chloropyrimidin-4-yl)amino)-1*H*-pyrazole-3-carboxylate (**18**). The title compound was prepared according to the general procedure 1, using methyl 3-amino-1*H*-pyrazole-5-carboxylate and 2,4-dichloropyrimidine. The mixture was stirred at 50 °C for 72 h to obtain the product (132 mg, 16%) as a white solid. 1H NMR (250 MHz, DMSO-d6) δ 13.70 (s, 1H), 10.63 (s, 1H), 8.20 (d, J = 5.9 Hz, 1H), 7.19–6.81 (m, 2H), 3.86 (s, 3H). 13C NMR (151 MHz, DMSO) δ 160.56, 159.39, 147.77, 133.08, 128.15, 127.41, 99.45, 52.04. MS-ESI *m/z* [M + Na]⁺: calcd 276.7, found 276.1.

methyl 5-((2,5-dichloropyrimidin-4-yl)amino)-1*H*-pyrazole-3-carboxylate (**19**). The title compound was prepared according to the general procedure 1, using methyl 3-amino-1*H*-pyrazole-5-carboxylate and 2,4,5-trichloropyrimidine. The mixture was stirred at 60 °C for 72 h to obtain the product (781 mg, 84%) as a beige solid. 1H NMR (250 MHz, DMSO-d6) δ 13.87 (s, 1H), 10.10 (s, 1H), 8.40 (s, 1H), 7.06 (s, 1H), 3.86 (s, 3H). 13C NMR (126 MHz, DMSO) δ 159.22, 156.95, 156.77, 155.60, 146.39, 133.11, 113.43, 102.23, 52.12. MS-ESI *m/z* [M + H]⁺: calcd 287.1, found 286.0.

methyl 5-((2-chloroquinazolin-4-yl)amino)-1*H*-pyrazole-3-carboxylate (**20**). The title compound was prepared according to the general procedure 1, using methyl 3-amino-1*H*-pyrazole-5-carboxylate and 2,4-dichloroquinazoline. The mixture was stirred at 60 °C for 72 h to obtain the product (823 mg, 84%) as a white solid. 1H NMR (300 MHz, DMSO-d6) δ 13.88 (s, 1H), 11.12 (s, 1H), 8.67 (d, J = 8.3 Hz, 1H), 7.88 (t, J = 7.6 Hz, 1H), 7.73 (d, J = 8.3 Hz, 1H), 7.61 (t, J = 7.6 Hz, 1H), 7.35 (s, 1H), 3.89 (s, 3H). 13C NMR (75 MHz, DMSO) δ 159.32, 158.58, 156.07, 150.82, 147.41, 134.16, 132.93, 126.87, 126.77, 123.63, 113.45, 101.83, 52.11. MS-ESI *m/z* [M + H]⁺: calcd 304.7, found 304.1.

methyl 5-((2-((4-(2-((*tert*-butoxycarbonyl)amino)ethyl)phenyl)amino)pyrimidin-4-yl)amino)-1*H*-pyrazole-3-carboxylate (**21a**). The title compound was prepared according to the general procedure 3, using **16** and *tert*-butyl 4-aminophenethylcarbamate. The mixture was stirred for 18 h under reflux to obtain the product (219 mg, 63%) as a light yellow solid. 1H NMR (250 MHz, DMSO-d6) δ 13.87 (s, 1H), 11.29 (s, 1H), 10.50 (s, 1H), 8.00 (d, J = 6.9 Hz, 1H), 7.47 (d, J = 8.0 Hz, 2H), 7.23 (d, J = 8.1 Hz, 2H), 7.06 (s, 1H), 6.90 (t, J = 5.3 Hz, 1H), 6.52 (s, 1H), 3.87 (s, 3H), 3.22–3.10 (m, 2H), 2.80–2.65 (m, 2H), 1.37 (s, 9H). 13C NMR (126 MHz, DMSO) δ 160.11, 155.54, 146.85, 136.35, 133.35, 129.22, 122.79, 100.87, 99.04, 77.52, 52.04, 41.52, 35.06, 28.24. MS-ESI *m/z* [M + H]⁺: calcd 454.5, found 454.1. HRMS *m/z* [M + Na]⁺: calcd 476.2017, found 476.2038. HPLC: t_R = 6.09, purity ≥ 95% (UV: 254/280 nm).

methyl 5-((2-((3-(2-((*tert*-butoxycarbonyl)amino)ethyl)phenyl)amino)pyrimidin-4-yl)amino)-1*H*-pyrazole-3-carboxylate (**21b**). The title compound was prepared according to the general procedure 3, using **16** and *tert*-butyl 3-aminophenethylcarbamate. The mixture was stirred for 18 h under reflux to obtain the product (178 mg, 49%) as a light yellow solid. 1H NMR (250 MHz, DMSO-d6) δ 13.85 (s, 1H), 11.33 (s, 1H), 10.54 (s, 1H), 8.01 (d, J = 6.7 Hz, 1H), 7.54–7.28 (m, 3H), 7.08 (d, J = 7.1 Hz, 1H), 6.98 (s, 1H), 6.85 (t, J = 5.6 Hz, 1H), 6.52 (s, 1H), 3.84 (s, 3H), 3.16–3.06 (m, 2H), 2.75–2.64 (m, 2H), 1.34 (s, 9H). 13C NMR (126 MHz, DMSO) δ 160.28, 159.49, 155.52, 153.45, 147.09, 140.65, 136.85, 133.51, 129.09, 125.61, 123.38, 121.02, 100.73, 99.19, 77.53, 51.93, 41.30, 35.37, 28.22. MS-ESI *m/z* [M + H]⁺: calcd 454.5, found 454.3. HRMS *m/z* [M + H]⁺: calcd 454.2197, found 454.2182. HPLC: t_R = 6.07, purity ≥ 95% (UV: 254/280 nm).

methyl 5-((2-((5-((*tert*-butoxycarbonyl)amino)pentyl)amino)pyrimidin-4-yl)amino)-1*H*-pyrazole-3-carboxylate (**21c**). The title compound was prepared according to the general procedure 4, using **16** and *tert*-butyl (5-aminopentyl)carbamate. The mixture was stirred for 5 h at 120 °C to obtain the product (142 mg, 45%) as a white solid. 1H NMR (400 MHz, DMSO-d6) δ 13.38 (s, 1H), 9.91 (d, J = 185.8 Hz, 1H), 7.83 (s, 1H), 7.21 (s, 1H), 6.75 (s, 2H), 6.04 (d, J = 53.4 Hz, 1H), 3.82 (s, 3H), 3.23 (q, J = 6.3 Hz, 2H), 2.91 (q, J = 6.4 Hz, 2H), 1.61–1.46 (m, 2H), 1.36 (s, 13H). 13C NMR (101 MHz, DMSO) δ 162.50, 160.04, 157.02, 156.06, 150.13, 142.02, 100.07, 96.24, 77.78, 52.27, 41.26, 40.66, 29.78, 29.39, 28.74, 24.33. MS-ESI *m/z* [M + H]⁺: calcd 420.5, found 420.9. HRMS *m/z* [M + H]⁺: calcd 420.2354, found 420.2350. HPLC: t_R = 5.57, purity ≥ 95% (UV: 254/280 nm).

methyl 5-((2-((2-(2-((*tert*-butoxycarbonyl)amino)ethoxy)ethyl)amino)pyrimidin-4-yl)amino)-1*H*-pyrazole-3-carboxylate (**21d**). The title compound was prepared according to the general procedure 4, using **16** and *tert*-butyl (2-(2-aminoethoxy)ethyl)carbamate. The mixture was stirred for 5 h at 120 °C to obtain the product (10 mg, 31%) as a colorless oil. 1H NMR (500 MHz, DMSO-d6) δ 13.39 (s, 1H), 9.91 (d, J = 245.4 Hz, 1H), 7.86 (s, 1H), 7.54–5.86 (m, 4H), 3.82 (s, 3H), 3.60–3.49 (m, 2H), 3.46–3.38 (m, 4H), 3.09 (q, J = 5.9 Hz, 2H), 1.36 (s, 9H). 13C NMR (126 MHz, DMSO) δ 162.14, 159.69, 156.87, 155.63, 149.28, 141.64, 132.78, 99.45, 96.01, 92.69, 77.62, 69.08, 51.72, 40.48, 28.22. MS-ESI *m/z* [M + H]⁺: calcd 422.5, found 422.9. HRMS *m/z* [M + H]⁺: calcd 422.2146, found 422.2135. HPLC: t_R = 5.18, purity ≥ 95% (UV: 254/280 nm).

methyl 5-((2-((6-((*tert*-butoxycarbonyl)amino)hexyl)amino)pyrimidin-4-yl)amino)-1*H*-pyrazole-3-carboxylate (**21e**). The title compound was prepared according to the general procedure 4, using **16** and *tert*-butyl (6-aminohexyl)carbamate. The mixture was stirred for 5 h at 120 °C to obtain the product (36 mg, 42%) as a colorless oil. 1H NMR (400 MHz, DMSO-d6) δ 13.37 (s, 1H), 9.90 (d, J = 169.7 Hz, 1H), 7.85 (s, 1H), 7.17 (s, 1H), 6.94–6.58 (m,

2H), 6.07 (s, 1H), 3.81 (s, 3H), 3.23 (q, J = 6.7 Hz, 2H), 2.98–2.83 (m, 2H), 1.57–1.45 (m, 2H), 1.40–1.23 (m, 15H). 13C NMR (101 MHz, DMSO) δ 162.04, 159.47, 156.64, 155.60, 149.06, 140.95, 132.46, 99.15, 95.72, 77.29, 51.72, 40.65, 29.53, 29.24, 28.27, 26.31, 26.18. MS-ESI *m/z* [M + H]⁺: calcd 434.5, found 435.0. HRMS *m/z* [M + H]⁺: calcd 434.2510, found 434.2500. HPLC: t_R = 5.86, purity ≥ 95% (UV: 254/280 nm).

methyl 5-((2-((4-(cyanomethyl)phenyl)amino)pyrimidin-4-yl)amino)-1*H*-pyrazole-3-carboxylate (**21f**). The title compound was prepared according to the general procedure 2, using 16 and 2-(4-aminophenyl)acetonitrile. The mixture was stirred for 18 h under reflux to obtain the product (46 mg, 61%) as a white solid. 1H NMR (300 MHz, DMSO-d6) δ 11.52 (s, 1H), 10.90 (s, 1H), 8.05 (d, J = 7.1 Hz, 1H), 7.58 (d, J = 8.1 Hz, 2H), 7.40 (d, J = 8.3 Hz, 2H), 7.01 (s, 1H), 6.56 (s, 1H), 4.08 (s, 2H), 3.88 (s, 3H). 13C NMR (75 MHz, DMSO) δ 160.20, 159.36, 152.75, 146.04, 143.59, 135.95, 133.88, 128.85, 127.98, 123.09, 119.23, 100.74, 99.31, 52.14, 21.99. MS-ESI *m/z* [M + H]⁺: calcd 350.4, found 350.2. HRMS *m/z* [M + H]⁺: calcd 350.1360, found 350.1363. HPLC: t_R = 6.79, purity ≥ 95% (UV: 254/280 nm).

methyl 5-((2-((4-(((*tert*-butoxycarbonyl)amino)methyl)benzyl)amino)pyrimidin-4-yl) amino)-1*H*-pyrazole-3-carboxylate (**21g**). The title compound was prepared according to the general procedure 4, using 16 and *tert*-butyl 4-(aminomethyl)benzylcarbamate. The mixture was stirred for 5 h at 120 °C to obtain the product (71 mg, 26%) as a white solid. 1H NMR (400 MHz, DMSO-d6) δ 13.39 (s, 1H), 9.95 (d, J = 203.0 Hz, 1H), 7.86 (s, 1H), 7.39–7.11 (m, 5H), 6.12 (s, 1H), 4.46 (d, J = 6.1 Hz, 2H), 4.08 (d, J = 6.1 Hz, 2H), 3.83 (s, 3H), 1.38 (s, 9H). 13C NMR (101 MHz, DMSO) δ 161.88, 159.45, 156.66, 155.78, 149.13, 141.49, 139.09, 138.33, 127.28, 126.96, 126.83, 96.28, 77.72, 51.72, 43.83, 43.15, 28.25. MS-ESI *m/z* [M + H]⁺: calcd 454.5, found 455.1 HRMS *m/z* [M + H]⁺: calcd 454.2197, found 454.2190. HPLC: t_R = 5.60, purity ≥ 95% (UV: 254/280 nm).

methyl 5-((2-((4-(((*tert*-butoxycarbonyl)amino)methyl)phenyl)amino)pyrimidin-4-yl) amino)-1*H*-pyrazole-3-carboxylate (**21h**). The title compound was prepared according to the general procedure 3, using 16 and *tert*-butyl 4-aminobenzylcarbamate. The mixture was stirred for 18 h under reflux to obtain the product (228 mg, 64%) as a white solid. 1H NMR (500 MHz, DMSO-d6) δ 13.18 (d, J = 722.0 Hz, 1H), 11.46 (s, 1H), 10.70 (s, 1H), 8.00 (s, 1H), 7.47 (d, J = 8.0 Hz, 2H), 7.41 (t, J = 6.2 Hz, 1H), 7.29 (d, J = 8.1 Hz, 2H), 7.21–6.97 (m, 1H), 6.50 (s, 1H), 4.16 (d, J = 6.0 Hz, 2H), 3.88 (s, 3H), 1.40 (s, 9H). 13C NMR (126 MHz, DMSO) δ 160.20, 155.83, 152.89, 143.48, 137.58, 135.01, 133.33, 127.77, 122.86, 100.87, 99.23, 77.86, 52.13, 43.05, 28.26. MS-ESI *m/z* [M + H]⁺: calcd 440.5, found 440.2 HRMS *m/z* [M + H]⁺: calcd 440.2041, found 440.2038. HPLC: t_R = 5.86, purity ≥ 95% (UV: 254/280 nm).

methyl 5-((2-((4-((*tert*-butoxycarbonyl)amino)benzyl)amino)pyrimidin-4-yl)amino)-1*H*-pyrazole-3-carboxylate (**21i**). The title compound was prepared according to the general procedure 4, using 16 and *tert*-butyl (4-(aminomethyl)phenyl)carbamate. The mixture was stirred for 3 h at 120 °C to obtain the product (45 mg, 17%) as a white solid. 1H NMR (400 MHz, DMSO-d6) δ 13.34 (s, 1H), 9.87 (d, J = 138.1 Hz, 1H), 9.24 (s, 1H), 7.87 (s, 1H), 7.59–6.81 (m, 6H), 6.09 (s, 1H), 4.41 (d, J = 6.1 Hz, 2H), 3.82 (s, 3H), 1.45 (s, 9H). 13C NMR (101 MHz, DMSO) δ 161.86, 159.46, 156.56, 152.80, 138.03, 134.23, 127.79, 127.51, 118.04, 96.14, 78.86, 51.71, 43.66, 28.14. MS-ESI *m/z* [M + H]⁺: calcd 440.5, found 441.0. HRMS *m/z* [M + H]⁺: calcd 440.2041, found 440.2040. HPLC: t_R = 5.79, purity ≥ 95% (UV: 254/280 nm).

methyl 5-((2-((4-(2-((*tert*-butoxycarbonyl)amino)ethyl)phenyl)amino)-5-chloropyrimid in-4-yl)amino)-1*H*-pyrazole-3-carboxylate (**22a**). The title compound was prepared according to the general procedure 3, using 17 and *tert*-butyl 4-aminophenethylcarbamate. The mixture was stirred for 18 h under reflux to obtain the product (14 mg, 5%) as a light yellow solid. 1H NMR (250 MHz, DMSO-d6) δ 9.92 (s, 1H), 9.88 (s, 1H), 8.23 (s, 1H), 7.50 (d, J = 7.4 Hz, 2H), 7.09 (d, J = 7.5 Hz, 2H), 7.00–6.77 (m, 2H), 3.85 (s, 3H), 3.18–3.01 (m, 2H), 2.72–2.58 (m, 2H), 1.36 (s, 9H). 13C NMR (75 MHz, DMSO) δ 160.79, 155.83, 155.40, 155.23, 151.51, 143.10, 137.10, 133.41, 128.67, 119.73, 103.59, 77.37, 51.73, 41.55, 34.90, 28.16. MS-ESI *m/z* [M + H]⁺: calcd 489.0, found 488.2. HRMS *m/z* [M + H]⁺: calcd 488.1808, found 488.1799. HPLC: t_R = 9.01, purity ≥ 95% (UV: 254/280 nm).

methyl 5-((2-((3-(2-((tert-butoxycarbonyl)amino)ethyl)phenyl)amino)-5-chloropyrimidin-4-yl)amino)-1H-pyrazole-3-carboxylate (22b). The title compound was prepared according to the general procedure 3, using 17 and tert-butyl 3-aminophenethylcarbamate. The mixture was stirred for 18 h under reflux to obtain the product (23 mg, 9%) as a light yellow solid. 1H NMR (300 MHz, DMSO-d6) δ 13.82–13.29 (m, 1H), 9.99 (s, 1H), 9.72 (s, 1H), 8.22 (s, 1H), 7.53 (d, J = 8.2 Hz, 1H), 7.42 (s, 1H), 7.17 (s, 1H), 6.88–6.73 (m, 2H), 6.62 (s, 1H), 3.83 (s, 3H), 3.22–3.04 (m, 2H), 2.70–2.56 (m, 2H), 1.37 (s, 9H). 13C NMR (75 MHz, DMSO) δ 157.65, 155.50, 154.62, 140.16, 139.83, 128.46, 126.57, 121.82, 118.74, 116.55, 103.68, 77.51, 51.59, 41.58, 35.77, 28.26. MS-ESI m/z [M + H]+: calcd 489.0, found 488.2. HRMS m/z [M + H]+: calcd 488.1808, found 488.1802. HPLC: t_R = 9.03, purity ≥ 95% (UV: 254/280 nm).

methyl 5-((2-((5-((tert-butoxycarbonyl)amino)pentyl)amino)-5-chloropyrimidin-4-yl)amino)-1H-pyrazole-3-carboxylate (22c). The title compound was prepared according to the general procedure 4, using 17 and tert-butyl (5-aminopentyl)carbamate. The mixture was stirred for 3 h at 80 °C to obtain the product (28 mg, 35%) as a white solid. 1H NMR (300 MHz, DMSO-d6) δ 13.50 (d, J = 68.2 Hz, 1H), 9.36 (d, J = 271.2 Hz, 1H), 7.99 (d, J = 26.2 Hz, 1H), 7.63–6.31 (m, 3H), 3.80 (s, 3H), 3.29–3.12 (m, 2H), 2.99–2.84 (m, 2H), 1.63–1.43 (m, 2H), 1.44–1.25 (m, 13H). 13C NMR (75 MHz, DMSO) δ 160.48, 160.03, 155.59, 155.20, 141.57, 140.72, 101.18, 95.10, 77.30, 51.49, 40.73, 29.27, 28.71, 28.25, 23.79. MS-ESI m/z [M + H]+: calcd 454.9, found 454.2. HRMS m/z [M + H]+: calcd 454.1964, found 454.1954. HPLC: t_R = 7.59, purity ≥ 95% (UV: 254/280 nm).

methyl 5-((2-((2-(2-((tert-butoxycarbonyl)amino)ethoxy)ethyl)amino)-5-chloropyrimidin-4-yl)amino)-1H-pyrazole-3-carboxylate (22d). The title compound was prepared according to the general procedure 4, using 17 and tert-butyl (2-(2-aminoethoxy)ethyl)carbamate. The mixture was stirred for 8 h at 80 °C to obtain the product (147 mg, 46%) as a white solid. 1H NMR (300 MHz, DMSO-d6) δ 13.53 (d, J = 61.3 Hz, 1H), 9.38 (d, J = 280.8 Hz, 1H), 8.02 (d, J = 22.6 Hz, 1H), 7.17 (d, J = 40.1 Hz, 1H), 6.77 (t, J = 5.5 Hz, 1H), 6.57 (s, 1H), 3.92–3.75 (m, 3H), 3.58–3.48 (m, 2H), 3.49–3.35 (m, 4H), 3.17–3.01 (m, 2H), 1.38 (s, 9H). 13C NMR (75 MHz, DMSO) δ 163.17, 160.84, 156.09, 155.59, 154.53, 142.03, 141.04, 101.91, 95.75, 78.10, 69.55, 69.41, 52.46, 51.95, 41.18, 28.69. MS-ESI m/z [M + H]+: calcd 456.9, found 456.3. HRMS m/z [M + H]+: calcd 457.1790, found 457.1785. HPLC: t_R = 7.35, purity ≥ 95% (UV: 254/280 nm).

methyl 5-((2-((6-((tert-butoxycarbonyl)amino)hexyl)amino)-5-chloropyrimidin-4-yl)amino)-1H-pyrazole-3-carboxylate (22e). The title compound was prepared according to the general procedure 4, using 17 and tert-butyl (6-aminohexyl)carbamate. The mixture was stirred for 8 h at 80 °C to obtain the product (177 mg, 54%) as a colorless oil. 1H NMR (300 MHz, DMSO-d6) δ 13.49 (d, J = 68.2 Hz, 1H), 9.33 (d, J = 284.3 Hz, 1H), 7.98 (d, J = 28.7 Hz, 1H), 7.17 (d, J = 43.4 Hz, 1H), 6.73 (t, J = 5.7 Hz, 1H), 6.65–6.41 (m, 1H), 3.90–3.76 (m, 3H), 3.29–3.14 (m, 2H), 2.95–2.83 (m, 2H), 1.60–1.44 (m, 2H), 1.39–1.20 (m, 15H). 13C NMR (75 MHz, DMSO) δ 162.58, 159.99, 155.54, 155.04, 153.78, 147.66, 132.63, 101.06, 95.10, 77.26, 51.91, 51.45, 40.72, 29.49, 29.00, 28.25, 26.23, 26.11. MS-ESI m/z [M + H]+: calcd 469.0, found 468.4. HRMS m/z [M + H]+: calcd 468.2121, found 468.2116. HPLC: t_R = 7.82, purity ≥ 95% (UV: 254/280 nm).

methyl 5-((2-((4-(2-((tert-butoxycarbonyl)amino)ethyl)phenyl)amino)quinazolin-4-yl)amino)-1H-pyrazole-3-carboxylate (23a). The title compound was prepared according to the general procedure 3, using 21 and tert-butyl 4-aminophenethylcarbamate. The mixture was stirred for 18 h at 70 °C to obtain the product (325 mg, 72%) as a yellow solid. 1H NMR (300 MHz, DMSO-d6) δ 13.32 (s, 1H), 11.84 (s, 1H), 10.76 (s, 1H), 8.73 (d, J = 8.2 Hz, 1H), 7.87 (t, J = 7.7 Hz, 1H), 7.61 (d, J = 8.2 Hz, 1H), 7.55–7.41 (m, 3H), 7.23 (d, J = 8.0 Hz, 2H), 7.13 (s, 1H), 6.92 (t, J = 5.4 Hz, 1H), 3.88 (s, 3H), 3.24–3.09 (m, 2H), 2.81–2.67 (m, 2H), 1.37 (s, 9H). 13C NMR (75 MHz, DMSO) δ 159.64, 158.44, 155.56, 151.92, 145.61, 139.54, 136.79, 135.78, 134.22, 129.28, 124.88, 124.78, 122.92, 117.59, 110.25, 102.96, 77.55, 52.06, 41.49, 35.13, 28.25. MS-ESI m/z [M + H]+: calcd 505.6, found 505.2. HRMS m/z [M + H]+: calcd 504.2354, found 504.2348. HPLC: t_R = 7.77, purity ≥ 95% (UV: 254/280 nm).

methyl 5-((2-((3-(2-((*tert*-butoxycarbonyl)amino)ethyl)phenyl)amino)quinazolin-4-yl)amino)-1*H*-pyrazole-3-carboxylate (**23b**). The title compound was prepared according to the general procedure 3, using **21** and *tert*-butyl 3-aminophenethylcarbamate. The mixture was stirred for 18 h at 70 °C to obtain the product (305 mg, 61%) as a yellow solid. 1H NMR (300 MHz, DMSO-d6) δ 13.65 (d, J = 246.4 Hz, 1H), 11.82 (s, 1H), 10.70 (s, 1H), 8.72 (d, J = 8.3 Hz, 1H), 7.88 (t, J = 7.8 Hz, 1H), 7.65 (d, J = 8.3 Hz, 1H), 7.50 (t, J = 7.7 Hz, 1H), 7.45–7.29 (m, 3H), 7.13 (d, J = 7.3 Hz, 1H), 7.06 (s, 1H), 6.85 (t, J = 5.9 Hz, 1H), 3.86 (s, 3H), 3.20–3.06 (m, 2H), 2.74–2.64 (m, 2H), 1.34 (s, 9H). 13C NMR (75 MHz, DMSO) δ 159.75, 158.31, 155.56, 152.14, 140.75, 139.90, 136.16, 135.74, 129.18, 126.19, 124.95, 124.70, 123.51, 121.24, 117.69, 110.35, 102.58, 77.60, 51.95, 41.28, 35.38, 28.24. MS-ESI *m/z* [M + H]+: calcd 505.6, found 505.4. HRMS *m/z* [M + H]+: calcd 504.2354, found 504.2347. HPLC: t_R = 8.11, purity ≥ 95% (UV: 254/280 nm).

methyl 5-((2-((5-((*tert*-butoxycarbonyl)amino)pentyl)amino)quinazolin-4-yl)amino)-1*H*-pyrazole-3-carboxylate (**23c**). The title compound was prepared according to the general procedure 4, using **21** and *tert*-butyl (5-aminopentyl)carbamate. The mixture was stirred for 8 h at 90 °C to obtain the product (233 mg, 50%) as a yellow solid. 1H NMR (250 MHz, DMSO-d6) δ 14.04 (s, 1H), 11.60 (s, 1H), 8.59 (d, J = 8.1 Hz, 1H), 8.10–7.65 (m, 2H), 7.62–7.24 (m, 2H), 6.76 (s, 1H), 3.87 (s, 3H), 3.53–3.42 (m, 2H), 3.03–2.87 (m, 2H), 1.75–1.54 (m, 2H), 1.46–1.25 (m, 13H). 13C NMR (75 MHz, DMSO) δ 156.10, 136.42, 135.13, 125.12, 124.67, 117.61, 110.68, 102.74, 77.82, 52.56, 41.71, 40.17, 29.66, 29.02, 28.73, 24.05. MS-ESI *m/z* [M + H]+: calcd 470.6, found 470.5. HRMS *m/z* [M + H]+: calcd 470.2510, found 470.2505. HPLC: t_R = 7.91, purity ≥ 95% (UV: 254/280 nm).

methyl 3-((2-((4-(2-aminoethyl)phenyl)amino)pyrimidin-4-yl)amino)-1*H*-pyrazole-5-carboxylate (**24**). **18a** (50 mg, 0.1 mmol) was dissolved in anhydrous DCM (4 mL). TFA (503 mg, 4.4 mmol) was added at 0°C and the reaction mixture was allowed to warm up to rt overnight. The solvent was evaporated under reduced pressure. The residue was dissolved in methanol and neutralized with saturated K_2CO_3 solution. The solvent was again evaporated under reduced pressure and the crude product was purified by flash chromatography using H_2O/acetonitrile as an eluent to obtain the desired product (21 mg, 54%) as a yellow solid. 1H NMR (250 MHz, DMSO-d6) δ 9.27 (s, 1H), 8.03 (d, J = 5.7 Hz, 1H), 7.69–7.53 (m, 2H), 7.19–7.06 (m, 2H), 6.81 (s, 1H), 6.30 (d, J = 5.7 Hz, 1H), 3.84 (s, 3H), 2.84–2.70 (m, 2H), 2.68–2.54 (m, 2H). 13C NMR (126 MHz, DMSO) δ 160.93, 159.55, 159.08, 156.32, 151.48, 145.39, 138.58, 132.90, 128.62, 119.03, 98.07, 96.95, 51.72, 43.72, 39.02. MS-ESI *m/z* [M + H]+: calcd 354.4, found 354.7. HRMS *m/z* [M + Na]+: calcd 376.1492, found 376.1485. HPLC: t_R = 10.09, purity ≥ 95% (UV: 254/280 nm).

methyl 3-((2-((4-aminobenzyl)amino)pyrimidin-4-yl)amino)-1*H*-pyrazole-5-carboxylate (**25**). **18i** (20 mg, 0.05 mmol) was dissolved in anhydrous DCM (1 mL). TFA (208 mg, 1.8 mmol) was added at 0 °C and the reaction mixture was allowed to warm up to rt overnight. The solvent was evaporated under reduced pressure. The residue was dissolved in methanol and neutralized with saturated K_2CO_3 solution. The solvent was again evaporated under reduced pressure and the crude product was purified by flash chromatography using H_2O/acetonitrile as an eluent to obtain the desired product (5 mg, 36%) as a yellow solid. 1H NMR (250 MHz, DMSO-d6) δ 13.38 (s, 1H), 9.81 (s, 1H), 7.88 (d, J = 5.5 Hz, 1H), 7.26 (s, 1H), 7.01 (d, J = 8.3 Hz, 2H), 6.50 (d, J = 8.3 Hz, 2H), 6.07 (d, J = 4.8 Hz, 1H), 4.89 (s, 2H), 4.30 (d, J = 5.9 Hz, 2H), 3.81 (s, 3H). MS-ESI *m/z* [M + H]+: calcd 340.4, found 340.6. HRMS *m/z* [M + H]+: calcd 340.1517, found 340.1529. HPLC: t_R = 10.13, purity ≥ 95% (UV: 254/280 nm).

2-chloro-*N*-(3-methyl-1*H*-pyrazol-5-yl)pyrimidin-4-amine (**32**). The title compound was prepared according to the general procedure 1, using 5-methyl-1*H*-pyrazol-3-amine and 2,4-dichloropyrimidine. The mixture was stirred at 60 °C for 72 h to obtain the product (466 mg, 48%) as a beige solid. 1H NMR (300 MHz, DMSO-d6) δ 12.13 (s, 1H), 10.28 (s, 1H), 8.15 (d, J = 5.9 Hz, 1H), 7.15 (s, 1H), 6.09 (s, 1H), 2.22 (s, 3H). 13C NMR (75 MHz, DMSO) δ 160.77, 159.37, 147.35, 142.17, 138.83, 104.97, 95.50, 10.61. MS-ESI *m/z* [M + H]+: calcd 210.6, found 210.2.

2-chloro-*N*-(3-isopropyl-1*H*-pyrazol-5-yl)pyrimidin-4-amine (**33**). The title compound was prepared according to the general procedure 1, using 5-isopropyl-1*H*-pyrazol-3-amine and 2,4-dichloropyrimidine. The mixture was stirred at 60 °C for 120 h to obtain the product (543 mg, 63%) as a beige solid. 1H NMR (300 MHz, DMSO-d6) δ 12.17 (s, 1H), 10.28 (s, 1H), 8.16 (d, J = 6.0 Hz, 1H), 7.20 (s, 1H), 6.07 (s, 1H), 2.93 (p, J = 6.9 Hz, 1H), 1.21 (d, J = 6.9 Hz, 6H). 13C NMR (75 MHz, DMSO) δ 161.09, 160.80, 159.35, 149.85, 147.06, 104.94, 92.75, 25.29, 22.21. MS-ESI *m/z* [M + H]$^+$: calcd 238.7, found 238.2.

N-(3-(*tert*-butyl)-1*H*-pyrazol-5-yl)-2-chloropyrimidin-4-amine (**34**). The title compound was prepared according to the general procedure 1, using 5-(*tert*-butyl)-1*H*-pyrazol-3-amine and 2,4-dichloropyrimidine. The mixture was stirred at 60 °C for 72 h to obtain the product (688 mg, 84%) as a beige solid. 1H NMR (300 MHz, DMSO-d6) δ 12.17 (s, 1H), 10.28 (s, 1H), 8.16 (d, J = 5.9 Hz, 1H), 7.24 (s, 1H), 6.04 (s, 1H), 1.26 (s, 9H). 13C NMR (75 MHz, DMSO) δ 160.80, 159.35, 157.60, 152.96, 146.72, 104.86, 92.35, 30.66, 29.89. MS-ESI *m/z* [M + H]$^+$: calcd 252.7, found 252.2.

5-((2-chloropyrimidin-4-yl)amino)-*N*-methyl-1*H*-pyrazole-3-carboxamide (**35**). The title compound was prepared according to the general procedure 1, using 3-amino-*N*-methyl-1*H*-pyrazole-5-carboxamide and 2,4-dichloropyrimidine. The mixture was stirred at 60 °C for 48 h to obtain the product (82 mg, 14%) as a white solid. 1H NMR (300 MHz, DMSO-d6) δ 13.18 (s, 1H), 10.49 (s, 1H), 8.52 (d, J = 5.2 Hz, 1H), 8.20 (d, J = 5.9 Hz, 1H), 7.18 (s, 1H), 6.85 (s, 1H), 2.76 (d, J = 4.5 Hz, 3H). 13C NMR (126 MHz, DMSO) δ 160.84, 160.77, 159.46, 158.95, 147.31, 137.29, 105.11, 95.89, 25.55. MS-ESI *m/z* [M + H]$^+$: calcd 253.7, found 253.2.

isopropyl 3-((2-chloropyrimidin-4-yl)amino)-1*H*-pyrazole-5-carboxylate (**36**). The title compound was prepared according to the general procedure 1, using isopropyl 3-amino-1*H*-pyrazole-5-carboxylate and 2,4-dichloropyrimidine. The mixture was stirred at 60 °C for 48 h to obtain the product (363 mg, 43%) as a beige solid with impurities. MS-ESI *m/z* [M + H]$^+$: calcd 282.7, found 282.1.

tert-butyl 3-((2-chloropyrimidin-4-yl)amino)-1*H*-pyrazole-5-carboxylate (**37**). The title compound was prepared according to the general procedure 1, using *tert*-butyl 3-amino-1*H*-pyrazole-5-carboxylate and 2,4-dichloropyrimidine. The mixture was stirred at 60 °C for 48 h to obtain the product (565 mg, 35%) as a white solid. 1H NMR (400 MHz, DMSO-d6) δ 13.49 (s, 1H), 10.61 (s, 1H), 8.20 (d, J = 5.9 Hz, 1H), 7.27–6.55 (m, 2H), 1.54 (s, 9H). 13C NMR (101 MHz, DMSO) δ 160.59, 159.40, 158.09, 147.59, 142.19, 134.61, 105.37, 99.20, 81.95, 27.79. MS-ESI *m/z* [M + H]$^+$: calcd 296.7, found 296.1.

tert-butyl 4-((4-((3-methyl-1*H*-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)phenethyl-carbamate (**38a**). The title compound was prepared according to the general procedure 2, using **28** and *tert*-butyl 4-aminophenethylcarbamate. The mixture was stirred for 18 h under reflux to obtain the product (66 mg, 67%) as a white solid. 1H NMR (300 MHz, DMSO-d6) δ 12.41 (s, 1H), 11.20 (s, 1H), 10.68 (s, 1H), 7.96 (s, 1H), 7.45 (d, J = 6.5 Hz, 2H), 7.25 (d, J = 8.0 Hz, 2H), 6.91 (t, J = 5.6 Hz, 1H), 6.45 (s, 1H), 6.24 (s, 1H), 3.16 (q, J = 6.9 Hz, 2H), 2.78–2.66 (m, 2H), 2.19 (s, 3H), 1.37 (s, 9H). 13C NMR (75 MHz, DMSO) δ 159.67, 155.54, 152.57, 145.99, 142.74, 139.27, 136.66, 134.69, 129.15, 123.31, 98.95, 97.03, 77.53, 41.57, 34.98, 28.25, 10.63. MS-ESI *m/z* [M + H]$^+$: calcd 410.5, found 410.4. HRMS *m/z* [M + H]$^+$: calcd 410.2299, found 410.2297. HPLC: t$_R$ = 7.30, purity ≥ 95% (UV: 254/280 nm).

tert-butyl (5-((4-((3-methyl-1*H*-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)pentyl)-carbamate (**38b**). The title compound was prepared according to the general procedure 4, using **28** and *tert*-butyl (5-aminopentyl)carbamate. The mixture was stirred for 8 h at 90 °C to obtain the product (31 mg, 29%) as a colorless oil. 1H NMR (300 MHz, DMSO-d6) δ 11.84 (s, 1H), 9.30 (s, 1H), 7.78 (d, J = 5.7 Hz, 1H), 6.75 (t, J = 5.7 Hz, 1H), 6.62 (s, 1H), 6.33–6.07 (m, 2H), 3.21 (q, J = 6.6 Hz, 2H), 2.91 (q, J = 6.5 Hz, 2H), 2.18 (s, 3H), 1.58–1.47 (m, 2H), 1.44–1.37 (m, 2H), 1.36 (s, 9H), 1.32–1.22 (m, 2H). 13C NMR (75 MHz, DMSO) δ 162.09, 159.66, 156.04, 155.60, 148.52, 138.40, 95.58, 95.15, 77.31, 41.30, 39.89, 29.36, 28.97, 28.28, 23.94, 11.07. MS-ESI *m/z* [M + H]$^+$: calcd 376.5, found 376.4. HRMS *m/z* [M + H]$^+$: calcd 376.2456, found 376.2458. HPLC: t$_R$ = 7.08, purity ≥ 95% (UV: 254/280 nm).

tert-butyl 4-((4-((3-isopropyl-1*H*-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)phenethyl-carbamate (**39a**). The title compound was prepared according to the general procedure 2, using **29** and *tert*-butyl 4-aminophenethylcarbamate. The mixture was stirred for 18 h at 70 °C to obtain the product (70 mg, 64%) as a white solid. 1H NMR (300 MHz, DMSO-d6) δ 12.41 (s, 1H), 11.22 (s, 1H), 10.62 (s, 1H), 7.96 (d, J = 5.8 Hz, 1H), 7.43 (d, J = 7.9 Hz, 2H), 7.24 (d, J = 8.2 Hz, 2H), 6.90 (t, J = 5.0 Hz, 1H), 6.45 (s, 1H), 6.22 (s, 1H), 3.21–3.07 (m, 2H), 2.98–2.81 (m, 1H), 2.74–2.66 (m, 2H), 1.37 (s, 9H), 1.15 (d, J = 6.9 Hz, 6H). 13C NMR (75 MHz, DMSO) δ 159.67, 155.52, 152.65, 149.96, 145.70, 142.19, 136.62, 134.50, 129.22, 123.45, 99.10, 94.05, 77.52, 41.45, 35.03, 28.25, 25.35, 22.21. MS-ESI *m/z* [M + H]⁺: calcd 438.6, found 438.5. HRMS *m/z* [M + H]⁺: calcd 438.2612, found 438.2609. HPLC: t_R = 7.92, purity ≥ 95% (UV: 254/280 nm).

tert-butyl (5-((4-((3-isopropyl-1*H*-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)pentyl)-carbamate (**39b**). The title compound was prepared according to the general procedure 4, using **29** and *tert*-butyl (5-aminopentyl)carbamate. The mixture was stirred for 8 h at 90 °C to obtain the product (22 mg, 22%) as a yellow solid. 1H NMR (300 MHz, DMSO-d6) δ 11.97 (s, 1H), 9.71 (s, 1H), 7.78 (d, J = 6.0 Hz, 1H), 6.98 (s, 1H), 6.74 (t, J = 5.7 Hz, 1H), 6.45–6.04 (m, 2H), 3.25 (q, J = 6.9 Hz, 2H), 2.96–2.82 (m, 3H), 1.62–1.47 (m, 2H), 1.45–1.26 (m, 13H), 1.21 (d, J = 6.9 Hz, 6H). 13C NMR (75 MHz, DMSO) δ 160.54, 159.70, 155.61, 153.36, 149.65, 147.25, 96.01, 92.67, 77.34, 40.75, 29.32, 28.86, 28.29, 25.61, 23.90, 22.34. MS-ESI *m/z* [M + H]⁺: calcd 404.5, found 404.5. HRMS *m/z* [M + H]⁺: calcd 404.2769, found 404.2767. HPLC: t_R = 7.81, purity ≥ 95% (UV: 254/280 nm).

tert-butyl 4-((4-((3-(*tert*-butyl)-1*H*-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)phenethyl-carbamate (**40a**). The title compound was prepared according to the general procedure 3, using **30** and *tert*-butyl 4-aminophenethylcarbamate. The mixture was stirred for 18 h under reflux to obtain the product (70 mg, 79%) as a yellow solid. 1H NMR (300 MHz, DMSO-d6) δ 12.38 (s, 1H), 11.24 (s, 1H), 10.65 (s, 1H), 7.96 (d, J = 4.5 Hz, 1H), 7.42 (d, J = 7.7 Hz, 2H), 7.23 (d, J = 8.0 Hz, 2H), 6.89 (t, J = 5.2 Hz, 1H), 6.45 (s, 1H), 6.21 (s, 1H), 3.22–3.08 (m, 2H), 2.76–2.62 (m, 2H), 1.37 (s, 9H), 1.20 (s, 9H). 13C NMR (75 MHz, DMSO) δ 159.59, 155.51, 152.98, 152.52, 145.46, 142.22, 136.67, 134.40, 129.26, 123.32, 99.15, 93.71, 77.51, 41.39, 35.02, 30.71, 29.87, 28.24. MS-ESI *m/z* [M + H]⁺: calcd 452.6, found 452.4. HRMS *m/z* [M + H]⁺: calcd 452.2769, found 452.2763. HPLC: t_R = 7.78, purity ≥ 95% (UV: 254/280 nm).

tert-butyl (5-((4-((3-(tert-butyl)-1*H*-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)pentyl)-carbamate (**40b**). The title compound was prepared according to the general procedure 4, using **30** and *tert*-butyl (5-aminopentyl)carbamate. The mixture was stirred for 8 h at 90 °C to obtain the product (24 mg, 24%) as a colorless oil. 1H NMR (300 MHz, DMSO-d6) δ 11.86 (s, 1H), 9.27 (s, 1H), 7.77 (d, J = 5.7 Hz, 1H), 6.73 (t, J = 5.3 Hz, 1H), 6.59 (s, 1H), 6.38 (s, 1H), 6.14 (s, 1H), 3.23 (q, J = 6.7 Hz, 2H), 2.90 (q, J = 6.5 Hz, 2H), 1.58–1.46 (m, 2H), 1.44–1.22 (m, 22H). 13C NMR (75 MHz, DMSO) δ 162.10, 159.74, 156.02, 155.57, 151.97, 148.06, 95.55, 92.47, 77.29, 40.69, 30.73, 30.05, 29.32, 29.03, 28.27, 23.92. MS-ESI *m/z* [M + H]⁺: calcd 418.6, found 418.5. HRMS *m/z* [M + H]⁺: calcd 418.2925, found 418.2924. HPLC: t_R = 7.65, purity ≥ 95% (UV: 254/280 nm).

2-(4-((4-((3-(*tert*-butyl)-1*H*-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)phenyl)acetonitrile (**40c**). The title compound was prepared according to the general procedure 3, using **30** and 2-(4-aminophenyl)acetonitrile. The mixture was stirred for 18 h under reflux to obtain the product (31 mg, 41%) as a white solid. 1H NMR (300 MHz, DMSO-d6) δ 12.43 (s, 1H), 11.26 (s, 1H), 10.81 (s, 1H), 7.99 (d, J = 6.9 Hz, 1H), 7.56 (d, J = 8.0 Hz, 2H), 7.39 (d, J = 8.3 Hz, 2H), 6.49 (s, 1H), 6.19 (s, 1H), 4.04 (s, 2H), 1.21 (s, 9H). 13C NMR (75 MHz, DMSO) δ 159.77, 153.14, 152.57, 145.39, 142.44, 136.06, 128.82, 127.96, 123.51, 119.02, 99.32, 93.71, 30.73, 29.90, 21.91. MS-ESI *m/z* [M + H]⁺: calcd 348.4, found 348.3. HRMS *m/z* [M + H]⁺: calcd 348.1931, found 348.1933. HPLC: t_R = 7.27, purity ≥ 95% (UV: 254/280 nm).

tert-butyl 4-((4-((3-(methylcarbamoyl)-1*H*-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)-phenethylcarbamate (**41a**). The title compound was prepared according to the general procedure 3, using **31** and *tert*-butyl 4-aminophenethylcarbamate. The mixture was stirred for 24 h at 70 °C to obtain the product (20 mg, 40%) as a yellow solid. 1H NMR (300 MHz,

DMSO-d6) δ 13.08 (s, 1H), 10.12 (d, J = 126.5 Hz, 1H), 9.40 (d, J = 98.9 Hz, 1H), 8.35 (s, 1H), 8.04 (s, 1H), 7.64 (d, J = 8.0 Hz, 2H), 7.10 (d, J = 8.1 Hz, 2H), 6.98–6.77 (m, 1H), 6.24 (s, 1H), 3.19–3.07 (m, 2H), 2.78 (d, J = 4.6 Hz, 3H), 2.68–2.59 (m, 2H), 1.37 (s, 9H). 13C NMR (75 MHz, DMSO) δ 164.36, 159.06, 155.55, 151.53, 142.19, 138.52, 132.28, 129.03, 128.71, 119.31, 114.78, 100.24, 98.02, 77.50, 41.76, 34.97, 28.28, 25.60. MS-ESI m/z [M + H]⁺: calcd 453.5, found 454.0. HRMS m/z [M + Na]⁺: calcd 475.2177, found 475.2173. HPLC: t_R = 7.04, purity ≥ 95% (UV: 254/280 nm).

tert-butyl (5-((4-((3-(methylcarbamoyl)-1*H*-pyrazol-5-yl)amino)pyrimidin-2-yl)amino)-pentyl)carbamate (**41b**). The title compound was prepared according to the general procedure 4, using **31** and *tert*-butyl (5-aminopentyl)carbamate. The mixture was stirred for 8 h at 90 °C to obtain the product (17 mg, 34%) as a white solid. 1H NMR (300 MHz, DMSO-d6) δ 12.95 (d, J = 48.6 Hz, 1H), 9.74 (d, J = 191.6 Hz, 1H), 8.36–7.75 (m, 2H), 7.33–6.86 (m, 1H), 6.75 (t, J = 4.9 Hz, 1H), 6.60–5.86 (m, 2H), 3.24 (q, J = 6.6 Hz, 2H), 2.90 (q, J = 6.5 Hz, 2H), 2.76 (d, J = 4.6 Hz, 3H), 1.59–1.45 (m, 2H), 1.45–1.21 (m, 13H). 13C NMR (75 MHz, DMSO) δ 164.36, 162.01, 159.74, 156.72, 155.62, 151.55, 142.21, 100.24, 95.70, 77.33, 40.58, 29.30, 28.93, 28.28, 25.59, 23.89. MS-ESI m/z [M + H]⁺: calcd 419.5, found 419.8. HRMS m/z [M + H]⁺: calcd 419.2514, found 419.2524. HPLC: t_R = 6.83, purity ≥ 95% (UV: 254/280 nm).

5-((2-((4-(cyanomethyl)phenyl)amino)pyrimidin-4-yl)amino)-N-methyl-1*H*-pyrazole-3-carboxamide (**41c**). The title compound was prepared according to the general procedure 2, using **31** and 2-(4-aminophenyl)acetonitrile. The mixture was stirred for 18 h under reflux to obtain the product (15 mg, 45%) as a yellow solid. 1H NMR (300 MHz, DMSO-d6) δ 11.28 (s, 1H), 10.90 (s, 1H), 8.56 (s, 1H), 8.14–7.92 (m, 1H), 7.61 (d, J = 8.1 Hz, 2H), 7.38 (d, J = 8.1 Hz, 2H), 7.19 (s, 1H), 6.57 (s, 1H), 4.04 (s, 2H), 2.79 (d, J = 4.5 Hz, 3H). 13C NMR (75 MHz, DMSO) δ 160.50, 159.22, 152.53, 145.57, 144.26, 138.51, 136.11, 128.96, 127.72, 122.60, 119.27, 99.11, 98.18, 25.60, 21.97. MS-ESI m/z [M + H]⁺: calcd 349.4, found 349.6. HRMS m/z [M + H]⁺: calcd 349.1520, found 349.1525. HPLC: t_R = 6.23, purity ≥ 95% (UV: 254/280 nm).

isopropyl 3-((2-((5-((*tert*-butoxycarbonyl)amino)pentyl)amino)pyrimidin-4-yl)amino)-1*H*-pyrazole-5-carboxylate (**42b**). The title compound was prepared according to the general procedure 4, using **32** and *tert*-butyl (5-aminopentyl)carbamate. The mixture was stirred for 6 h at 100 °C to obtain the product (8 mg, 4%) as a yellow solid. 1H NMR (400 MHz, DMSO-d6) δ 13.32 (s, 1H), 9.89 (d, J = 172.8 Hz, 1H), 7.83 (s, 1H), 7.15 (s, 1H), 6.74 (s, 2H), 6.10 (s, 1H), 5.11 (q, J = 12.9, 6.6 Hz, 1H), 3.24 (q, J = 6.6 Hz, 2H), 2.90 (q, J = 6.6 Hz, 2H), 1.58–1.47 (m, 2H), 1.44–1.28 (m, 19H). 13C NMR (101 MHz, DMSO) δ 156.59, 155.58, 96.06, 91.63, 77.29, 68.23, 29.29, 28.88, 28.26, 23.86, 21.66, 13.39, 12.99. MS-ESI m/z [M + H]⁺: calcd 448.5, found 448.5. HRMS m/z [M + H]⁺: calcd 448.2673, found 448.2667. HPLC: t_R = 6.16, purity ≥ 95% (UV: 254/280 nm).

isopropyl 3-((2-((4-((*tert*-butoxycarbonyl)amino)benzyl)amino)pyrimidin-4-yl)amino)-1*H*-pyrazole-5-carboxylate (**42d**). The title compound was prepared according to the general procedure 4, using **32** and *tert*-butyl (4-(aminomethyl)phenyl)carbamate. The mixture was stirred for 6 h at 100 °C to obtain the product (8 mg, 4%) as a beige solid. 1H NMR (400 MHz, DMSO-d6) δ 13.31 (s, 1H), 9.91 (d, J = 194.0 Hz, 1H), 9.23 (s, 1H), 7.83 (s, 1H), 7.65–6.93 (m, 6H), 6.12 (s, 1H), 5.11 (s, 1H), 4.42 (d, J = 6.2 Hz, 2H), 1.45 (s, 9H), 1.30 (d, J = 6.3 Hz, 6H). 13C NMR (101 MHz, DMSO) δ 162.00, 159.59, 158.67, 156.43, 152.78, 148.97, 137.95, 134.39, 127.73, 127.25, 117.99, 99.18, 96.05, 78.84, 68.27, 43.59, 28.13, 21.65. MS-ESI m/z [M + H]⁺: calcd 468.5, found 468.5. HRMS m/z [M + H]⁺: calcd 468.2351, found 468.2354. HPLC: t_R = 6.24, purity ≥ 95% (UV: 254/280 nm).

tert-butyl 3-((2-((5-((*tert*-butoxycarbonyl)amino)pentyl)amino)pyrimidin-4-yl)amino)-1*H*-pyrazole-5-carboxylate (**43b**). The title compound was prepared according to the general procedure 4, using **40** and *tert*-butyl (5-aminopentyl)carbamate. The mixture was stirred for 6 h at 100 °C to obtain the product (14 mg, 5%) as a yellow oil. 1H NMR (400 MHz, DMSO-d6) δ 13.25 (d, J = 59.3 Hz, 1H), 9.86 (d, J = 176.8 Hz, 1H), 7.81 (s, 1H), 7.10 (s, 1H), 6.73 (s, 2H), 6.10 (s, 1H), 3.24 (q, J = 6.6 Hz, 2H), 2.90 (q, J = 6.6 Hz, 2H), 1.56–1.47 (m, 11H), 1.42–1.26 (m, 13H). 13C NMR (101 MHz, DMSO) δ 162.18, 159.53, 158.61, 156.31, 155.57,

149.02, 134.08, 99.07, 95.80, 81.44, 77.29, 40.76, 29.31, 28.88, 28.27, 28.13, 27.85, 23.88. MS-ESI m/z [M + H]$^+$: calcd 462.6, found 462.5. HRMS m/z [M + H]$^+$: calcd 462.2819, found 462.2823. HPLC: t_R = 6.36, purity $\geq$ 95% (UV: 254/280 nm).

tert-butyl 3-((2-((4-((*tert*-butoxycarbonyl)amino)benzyl)amino)pyrimidin-4-yl)amino)-1*H*-pyrazole-5-carboxylate (**43d**). The title compound was prepared according to the general procedure 4, using **40** and *tert*-butyl (4-(aminomethyl)phenyl)carbamate. The mixture was stirred for 6 h at 100 °C to obtain the product (26 mg, 12%) as a white solid. 1H NMR (400 MHz, DMSO-d6) δ 13.26 (d, J = 60.0 Hz, 1H), 10.26–9.51 (m, 1H), 9.39–9.05 (m, 1H), 7.83 (s, 1H), 7.65–7.02 (m, 6H), 6.13 (s, 1H), 4.42 (d, J = 6.2 Hz, 2H), 1.52 (s, 9H), 1.45 (s, 9H). 13C NMR (101 MHz, DMSO) δ 162.05, 159.57, 158.46, 156.39, 152.79, 148.89, 137.99, 134.44, 127.79, 127.38, 117.99, 99.26, 96.07, 81.51, 78.85, 43.63, 28.14, 27.85. MS-ESI m/z [M + H]$^+$: calcd 482.6, found 482.5. HRMS m/z [M + H]$^+$: calcd 482.2507, found 482.2510. HPLC: t_R = 6.58, purity $\geq$ 95% (UV: 254/280 nm).

Abbreviations

CDK: cyclin-dependent kinase, DCM, dichloromethane; DSF, differential scanning fluorimetry; EE, ethyl acetate; FUCCI, florescent ubiquitination-base cell cycle indicator; HATU, hexafluorophosphate azabenzotriazole tetramethyl uronium; n.d., not determined; NSCLC, non-small-cell lung carcinoma; oN, overnight; rt, room temperature; SCC, squamous cell carcinoma; TEA, triethylamine; TFA, trifluoroacetic acid.

References

1. Wells, C.I.; Vasta, J.D.; Corona, C.R.; Wilkinson, J.; Zimprich, C.A.; Ingold, M.R.; Pickett, J.E.; Drewry, D.H.; Pugh, K.M.; Schwinn, M.K.; et al. Quantifying CDK Inhibitor Selectivity in Live Cells. *Nat. Commun.* **2020**, *11*, 2743. [CrossRef] [PubMed]
2. Lim, S.; Kaldis, P. Cdks, Cyclins and CKIs: Roles beyond Cell Cycle Regulation. *Development* **2013**, *140*, 3079–3093. [CrossRef] [PubMed]
3. Endicott, J.A.; Noble, M.E.M. Structural Characterization of the Cyclin-Dependent Protein Kinase Family. *Biochem Soc. Trans.* **2013**, *41*, 1008–1016. [CrossRef] [PubMed]
4. Sridhar, J.; Akula, N.; Pattabiraman, N. Selectivity and Potency of Cyclin-Dependent Kinase Inhibitors. *AAPS J.* **2006**, *8*, 204–221. [CrossRef] [PubMed]

5. Jeronimo, C.; Collin, P.; Robert, F. The RNA Polymerase II CTD: The Increasing Complexity of a Low-Complexity Protein Domain. *J. Mol. Biol.* 2016, 428, 2607–2622. [CrossRef] [PubMed]

6. Wood, D.J.; Endicott, J.A. Structural Insights into the Functional Diversity of the CDK–Cyclin Family. *Open Biol.* 2018, 8, 180112. [CrossRef]

7. Noble, M.; Barrett, P.; Endicott, J.; Johnson, L.; McDonnell, J.; Robertson, G.; Zawaira, A. Exploiting Structural Principles to Design Cyclin-Dependent Kinase Inhibitors. *Biochim. Biophys. Acta - Proteins Proteomics* 2005, 1754, 58–64. [CrossRef]

8. Attwood, M.M.; Fabbro, D.; Sokolov, A.V.; Knapp, S.; Schiöth, H.B. Trends in Kinase Drug Discovery: Targets, Indications and Inhibitor Design. *Nat. Rev. Drug Discov.* 2021, 20, 839–861. [CrossRef] [PubMed]

9. Goel, S.; Bergholz, J.S.; Zhao, J.J. Targeting CDK4 and CDK6 in Cancer. *Nat. Rev. Cancer* 2022, 22, 356–372. [CrossRef]

10. Suryadinata, R.; Sadowski, M.; Sarcevic, B. Control of Cell Cycle Progression by Phosphorylation of Cyclin-Dependent Kinase (CDK) Substrates. *Biosci. Rep.* 2010, 30, 243–255. [CrossRef]

11. Davidson, G.; Niehrs, C. Emerging Links between CDK Cell Cycle Regulators and Wnt Signaling. *Trends Cell Biol.* 2010, 20, 453–460. [CrossRef] [PubMed]

12. Huang, C.; Du, R.; Jia, X.; Liu, K.; Qiao, Y.; Wu, Q.; Yao, N.; Yang, L.; Zhou, L.; Liu, X.; et al. CDK15 Promotes Colorectal Cancer Progression via Phosphorylating PAK4 and Regulating β-Catenin/ MEK-ERK Signaling Pathway. *Cell Death Differ.* 2022, 29, 14–27. [CrossRef] [PubMed]

13. Malumbres, M.; Harlow, E.; Hunt, T.; Hunter, T.; Lahti, J.M.; Manning, G.; Morgan, D.O.; Tsai, L.H.; Wolgemuth, D.J. Cyclin-Dependent Kinases: A Family Portrait. *Nat. Cell Biol.* 2009, 11, 1275–1276. [CrossRef] [PubMed]

14. Liu, M.; Xu, Z.; Du, Z.; Wu, B.; Jin, T.; Xu, K.; Xu, L.; Li, E.; Xu, H. The Identification of Key Genes and Pathways in Glioma by Bioinformatics Analysis. *J. Immunol. Res.* 2017, 2017, 1278081. [CrossRef] [PubMed]

15. de Oliveira Pepino, R.; Coelho, F.; Janku, T.A.B.; Alencar, D.P.; de Azevedo, W.F.; Canduri, F. Overview of PCTK3/CDK18: A Cyclin-Dependent Kinase Involved in Specific Functions in Post-Mitotic Cells. *Curr. Med. Chem.* 2021, 28, 6846–6865. [CrossRef]

16. Malumbres, M. Cyclin-Dependent Kinases. *Genome Biol.* 2014, 15, 122. [CrossRef]

17. Mikolcevic, P.; Sigl, R.; Rauch, V.; Hess, M.W.; Pfaller, K.; Barisic, M.; Pelliniemi, L.J.; Boesl, M.; Geley, S. Cyclin-Dependent Kinase 16/PCTAIRE Kinase 1 Is Activated by Cyclin Y and Is Essential for Spermatogenesis. *Mol. Cell. Biol.* 2012, 32, 868–879. [CrossRef]

18. Shehata, S.N.; Deak, M.; Collodet, C.; Spiegl, S.; Geley, S.; Sumpton, D.; Sakamoto, K. Identification of Novel PCTAIRE-1/CDK16 Substrates Using a Chemical Genetic Screen. *Cell. Signal.* 2019, 59, 53–61. [CrossRef]

19. Le Bouffant, F.; Le Minter, P.; Traiffort, E.; Ruat, M.; Sladeczek, F. Multiple Subcellular Localizations of PCTAIRE-1 in Brain. *Mol. Cell. Neurosci.* 2000, 16, 388–395. [CrossRef]

20. Besset, V.; Rhee, K.; Wolgemuth, D.J. The Cellular Distribution and Kinase Activity of the Cdk Family Member Pctaire 1 in the Adult Mouse Brain and Testis Suggest Functions in Differentiation. *Cell Growth Differ.* 1999, 10, 173–181.

21. Chen, X.Y.; Gu, X.T.; Saiyin, H.; Wan, B.; Zhang, Y.J.; Li, J.; Wang, Y.L.; Gao, R.; Wang, Y.F.; Dong, W.P.; et al. Brain-Selective Kinase 2 (BRSK2) Phosphorylation on PCTAIRE1 Negatively Regulates Glucose-Stimulated Insulin Secretion in Pancreatic β-Cells. *J. Biol. Chem.* 2012, 287, 30368–30375. [CrossRef] [PubMed]

22. Palmer, K.J.; Konkel, J.E.; Stephens, D.J. PCTAIRE Protein Kinases Interact Directly with the COPII Complex and Modulate Secretory Cargo Transport. *J. Cell Sci.* 2005, 118, 3839–3847. [CrossRef] [PubMed]

23. Shimizu, K.; Uematsu, A.; Imai, Y.; Sawasaki, T. Pctaire1/Cdk16 Promotes Skeletal Myogenesis by Inducing Myoblast Migration and Fusion. *FEBS Lett.* 2014, 588, 3030–3037. [CrossRef] [PubMed]

24. Whittaker, S.R.; Mallinger, A.; Workman, P.; Clarke, P.A. Inhibitors of Cyclin-Dependent Kinases as Cancer Therapeutics. *Pharmacol. Ther.* 2017, 173, 83–105. [CrossRef] [PubMed]

25. Charrasse, S.; Carena, I.; Hagmann, J.; Woods-Cook, K.; Ferrari, S. PCTAIRE-1: Characterization, Subcellular Distribution, and Cell Cycle- Dependent Kinase Activity. *Cell Growth Differ.* 1999, 10, 611–620.

26. Mikolcevic, P.; Rainer, J.; Geley, S. Orphan Kinases Turn Eccentric. A New Class of Cyclin Y-Activated, Membrane-Targeted CDKs. *Cell Cycle* 2012, 11, 3758–3768. [CrossRef] [PubMed]

27. Yanagi, T.; Matsuzawa, S.I. PCTAIRE1/PCTK1/CDK16: A New Oncotarget? *Cell Cycle* 2015, 14, 463–464. [CrossRef]

28. Yanagi, T.; Hata, H.; Mizuno, E.; Kitamura, S.; Imafuku, K.; Nakazato, S.; Wang, L.; Nishihara, H.; Tanaka, S.; Shimizu, H. PCTAIRE1/CDK16/PCTK1 Is Overexpressed in Cutaneous Squamous Cell Carcinoma and Regulates P27 Stability and Cell Cycle. *J. Dermatol. Sci.* 2017, 86, 149–157. [CrossRef]

29. Yanagi, T.; Imafuku, K.; Kitamura, S.; Hata, H.; Shimizu, H. CDK16/PCTK1/PCTAIRE1 Is Highly Expressed in Melanomas but Not in Melanocytic Nevi or Sarcomas. *J. Dermatol.* 2019, 46, 634–636. [CrossRef]

30. Chu, I.M.; Hengst, L.; Slingerland, J.M. The Cdk Inhibitor P27 in Human Cancer: Prognostic Potential and Relevance to Anticancer Therapy. *Nat. Rev. Cancer* 2008, 8, 253–267. [CrossRef]

31. Wang, H.; Liu, H.; Min, S.; Shen, Y.; Li, W.; Chen, Y.; Wang, X. CDK16 Overexpressed in Non-Small Cell Lung Cancer and Regulates Cancer Cell Growth and Apoptosis via a P27-Dependent Mechanism. *Biomed. Pharmacother.* 2018, 103, 399–405. [CrossRef] [PubMed]

32. Xie, J.; Li, Y.; Jiang, K.; Hu, K.; Zhang, S.; Dong, X.; Dai, X.; Liu, L.; Zhang, T.; Yang, K.; et al. CDK16 Phosphorylates and Degrades P53 to Promote Radioresistance and Predicts Prognosis in Lung Cancer. *Theranostics* 2018, 8, 650–662. [CrossRef] [PubMed]

33. Vervoorts, J.; Neumann, D.; Lüscher, B. The CCNY (Cyclin Y)-CDK16 Kinase Complex: A New Regulator of Autophagy Downstream of AMPK. *Autophagy* 2020, 16, 1724–1726. [CrossRef] [PubMed]

34. Dohmen, M.; Krieg, S.; Agalaridis, G.; Zhu, X.; Shehata, S.N.; Pfeiffenberger, E.; Amelang, J.; Bütepage, M.; Buerova, E.; Pfaff, C.M.; et al. AMPK-Dependent Activation of the Cyclin Y/CDK16 Complex Controls Autophagy. *Nat. Commun.* **2020**, *11*, 1032. [CrossRef] [PubMed]

35. Axtman, A.; Drwry, D.; Wells, C. CDK16: The Pick of the Understudied PCTAIRE Kinases. *Nat. Rev. Drug Discov.* **2019**, *18*, 2019. [CrossRef]

36. Jhaveri, K.; Burris, H.A.; Yap, T.A.; Hamilton, E.; Rugo, H.S.; Goldman, J.W.; Dann, S.; Liu, F.; Wong, G.Y.; Krupka, H.; et al. The Evolution of Cyclin Dependent Kinase Inhibitors in the Treatment of Cancer. *Expert Rev. Anticancer Ther.* **2021**, *21*, 1105–1124. [CrossRef]

37. Zhang, M.; Zhang, L.; Hei, R.; Li, X.; Cai, H.; Wu, X.; Zheng, Q.; Cai, C. CDK Inhibitors in Cancer Therapy, an Overview of Recent Development. *Am. J. Cancer Res.* **2021**, *11*, 1913–1935.

38. Ferguson, F.M.; Doctor, Z.M.; Ficarro, S.B.; Browne, C.M.; Marto, J.A.; Johnson, J.L.; Yaron, T.M.; Cantley, L.C.; Kim, N.D.; Sim, T.; et al. Discovery of Covalent CDK14 Inhibitors with Pan-TAIRE Family Specificity. *Cell Chem. Biol.* **2019**, *26*, 804–817.e12. [CrossRef]

39. Antolin, A.A.; Ameratunga, M.; Banerji, U.; Clarke, P.A.; Workman, P.; Al-Lazikani, B. The Kinase Polypharmacology Landscape of Clinical PARP Inhibitors. *Sci. Rep.* **2020**, *10*, 2585. [CrossRef]

40. Dixon-Clarke, S.E.; Shehata, S.N.; Krojer, T.; Sharpe, T.D.; Von Delft, F.; Sakamoto, K.; Bullock, A.N. Structure and Inhibitor Specificity of the PCTAIRE-Family Kinase CDK16. *Biochem. J.* **2017**, *474*, 699–713. [CrossRef]

41. Statsuk, A.V.; Maly, D.J.; Seeliger, M.A.; Fabian, M.A.; Biggs, W.H.; Lockhart, D.J.; Zarrinkar, P.P.; Kuriyan, J.; Shokat, K.M. Tuning a Three-Component Reaction for Trapping Kinase Substrate Complexes. *J. Am. Chem. Soc.* **2008**, *130*, 17568–17574. [CrossRef] [PubMed]

42. El-gamal, M.I.; Zaraei, S.O.; Madkour, M.M.; Anbar, H.S. Evaluation of Substituted Pyrazole-Based Kinase Inhibitors in One Decade (2011–2020): Current Status and Future Prospects. *Molecules* **2022**, *27*, 330. [CrossRef] [PubMed]

43. Nitulescu, G.M. Quantitative and Qualitative Analysis of the Anti-Proliferative Potential of the Pyrazole Scaffold in the Design of Anticancer Agents. *Molecules* **2022**, *27*, 8854. [CrossRef]

44. Carter, T.A.; Wodicka, L.M.; Shah, N.P.; Velasco, A.M.; Fabian, M.A.; Treiber, D.K.; Milanov, Z.V.; Atteridge, C.E.; Biggs, W.H.; Edeen, P.T.; et al. Inhibition of Drug-Resistant Mutants of ABL, KIT, and EGF Receptor Kinases. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 11011–11016. [CrossRef]

45. Yu, C.; Mannan, A.M.; Yvone, G.M.; Ross, K.N.; Zhang, Y.-L.; Marton, M.A.; Taylor, B.R.; Crenshaw, A.; Gould, J.Z.; Tamayo, P.; et al. High-Throughput Identification of Genotype-Specific Cancer Vulnerabilities in Mixtures of Barcoded Tumor Cell Lines. *Nat. Biotechnol.* **2016**, *34*, 419–423. [CrossRef]

46. Fedorov, O.; Niesen, F.H.; Knapp, S. Kinase Inhibitor Selectivity Profiling Using Differential Scanning Fluorimetry. *Methods Mol. Biol.* **2012**, *795*, 109–118. [PubMed]

47. Karaman, M.W.; Herrgard, S.; Treiber, D.K.; Gallant, P.; Atteridge, C.E.; Campbell, B.T.; Chan, K.W.; Ciceri, P.; Davis, M.I.; Edeen, P.T.; et al. A Quantitative Analysis of Kinase Inhibitor Selectivity. *Nat. Biotechnol.* **2008**, *26*, 127–132. [CrossRef]

48. Zhang, S.; Jin, S.; Griffin, C.; Feng, Z.; Lin, J.; Venkatakrishnan, K.; Gupta, N. Effects of Itraconazole and Rifampin on the Pharmacokinetics of Mobocertinib (TAK-788), an Oral Epidermal Growth Factor Receptor Inhibitor, in Healthy Volunteers. *Clin. Pharmacol. Drug Dev.* **2021**, *10*, 1044–1053. [CrossRef]

49. Li, F.; MacKenzie, K.R.; Jain, P.; Santini, C.; Young, D.W.; Matzuk, M.M. Metabolism of JQ1, an Inhibitor of Bromodomain and Extra Terminal Bromodomain Proteins, in Human and Mouse Liver Microsomes. *Biol. Reprod.* **2020**, *103*, 427–436. [CrossRef]

50. Hanke, T.; Dehm, F.; Liening, S.; Popella, S.D.; MacZewsky, J.; Pillong, M.; Kunze, J.; Weinigel, C.; Barz, D.; Kaiser, A.; et al. Aminothiazole-Featured Pirinixic Acid Derivatives as Dual 5-Lipoxygenase and Microsomal Prostaglandin E2 Synthase-1 Inhibitors with Improved Potency and Efficiency in Vivo. *J. Med. Chem.* **2013**, *56*, 9031–9044. [CrossRef]

51. Yano, S.; Hoffman, R.M. Real-Time Determination of the Cell-Cycle Position of Individual Cells within Live Tumors Using FUCCI Cell-Cycle Imaging Shuya. *Cells* **2018**, *7*, 168. [CrossRef] [PubMed]

52. Zielke, N.; Edgar, B.A. FUCCI Sensors: Powerful New Tools for Analysis of Cell Proliferation. *Wiley Interdiscip. Rev. Dev. Biol.* **2015**, *4*, 469–487. [CrossRef]

53. Tjaden, A.; Chaikuad, A.; Kowarz, E.; Marschalek, R.; Knapp, S.; Schröder, M.; Müller, S. Image-Based Annotation of Chemogenomic Libraries for Phenotypic Screening. *Molecules* **2022**, *27*, 1439. [CrossRef] [PubMed]

54. Kalinichenko, S.G.; Matveeva, N.Y. Morphological Characteristics of Apoptosis and Its Significance in Neurogenesis. *Neurosci. Behav. Physiol.* **2008**, *38*, 333–344. [CrossRef] [PubMed]

55. Brasca, M.G.; Amboldi, N.; Ballinari, D.; Cameron, A.; Casale, E.; Cervi, G.; Colombo, M.; Colotta, F.; Croci, V.; D'Alessio, R.; et al. Identification of N,1,4,4-Tetramethyl-8-[[4-(4-Methylpiperazin-1-Yl)Phenyl] Amino]-4,5-Dihydro-1H-Pyrazolo[4,3-h]Quinazoline-3-Carboxamide (PHA-848125), a Potent, Orally Available Cyclin Dependent Kinase Inhibitor. *J. Med. Chem.* **2009**, *52*, 5152–5163. [CrossRef]

56. Krämer, A.; Kurz, C.G.; Berger, B.T.; Celik, I.E.; Tjaden, A.; Greco, F.A.; Knapp, S.; Hanke, T. Optimization of Pyrazolo[1,5-a]Pyrimidines Lead to the Identification of a Highly Selective Casein Kinase 2 Inhibitor. *Eur. J. Med. Chem.* **2020**, *208*, 112770. [CrossRef]

57. Chernobrovkin, A.L.; Cázares-Körner, C.; Friman, T.; Caballero, I.M.; Amadio, D.; Martinez Molina, D. A Tale of Two Tails: Efficient Profiling of Protein Degraders by Specific Functional and Target Engagement Readouts. *SLAS Discov.* **2021**, *26*, 534–546. [CrossRef]

58. Vasta, J.D.; Corona, C.R.; Wilkinson, J.; Zimprich, C.A.; Hartnett, J.R.; Ingold, M.R.; Zimmerman, K.; Machleidt, T.; Kirkland, T.A.; Huwiler, K.G.; et al. Quantitative, Wide-Spectrum Kinase Profiling in Live Cells for Assessing the Effect of Cellular ATP on Target Engagement. *Cell Chem. Biol.* **2018**, *25*, 206–214.e11. [CrossRef]

59. Kowarz, E.; Löscher, D.; Marschalek, R. Optimized Sleeping Beauty Transposons Rapidly Generate Stable Transgenic Cell Lines. *Biotechnol. J.* **2015**, *10*, 647–653. [CrossRef]

Appendix I: A Machine Learning and Live-Cell Imaging Tool Kit Uncovers Small Molecules Induced Phospholipidosis

Contribution Menge, A.: Conceptualization, Methodology, Investigation, Validation and Formal Analysis, writing – Original Draft, writing, review and editing.

Cell Chemical Biology

A machine learning and live-cell imaging tool kit uncovers small molecules induced phospholipidosis

Graphical abstract
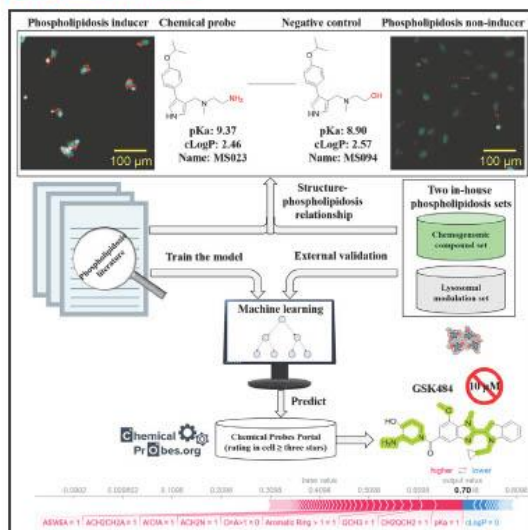
Authors

Huabin Hu, Amelie Tjaden, Stefan Knapp, Albert A. Antolin, Susanne Müller

Correspondence

aantolin@idibell.cat (A.A.A.), susanne.mueller-knapp@bmls.de (S.M.)

In brief

Hu et al. present an improved machine learning method as well as live-cell assay to detect drug-induced phospholipidosis, a pathological condition induced by pharmacological substances. Applying these methods, the differential phospholipidosis-inducing capacities of closely related molecules, such as chemical probes and matched negative control compound were identified.

Highlights

- Drug-induced phospholipidosis (DIPL) can lead to clinical adverse effects

- A live-cell imaging and machine learning tool kit for DIPL detection is presented

- Closely related molecules (chemical probes and controls) induce DIPL differentially

CellPress

Cell Chemical Biology

CellPress

Article

A machine learning and live-cell imaging tool kit uncovers small molecules induced phospholipidosis

Huabin Hu,[1,2,6] Amelie Tjaden,[3,4,6] Stefan Knapp,[3,4] Albert A. Antolin,[1,5,*] and Susanne Müller[3,4,7,*]

[1]Centre for Cancer Drug Discovery, Division of Cancer Therapeutics, The Institute of Cancer Research, London, UK
[2]Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, BMC, Box 596, SE-751 24 Uppsala, Sweden
[3]Institute of Pharmaceutical Chemistry, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany
[4]Structural Genomics Consortium (SGC), Buchmann Institute for Life Sciences, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany
[5]ProCURE, Catalan Institute of Oncology, Oncobell, Bellvitge Institute for Biomedical Research (IDIBELL), L'Hospitalet del Llobregat, Catalonia Barcelona, Spain
[6]These authors contributed equally
[7]Lead contact
*Correspondence: aantolin@idibell.cat (A.A.A.), susanne.mueller-knapp@bmls.de (S.M.)
https://doi.org/10.1016/j.chembiol.2023.09.003

SUMMARY

Drug-induced phospholipidosis (DIPL), characterized by excessive accumulation of phospholipids in lysosomes, can lead to clinical adverse effects. It may also alter phenotypic responses in functional studies using chemical probes. Therefore, robust methods are needed to predict and quantify phospholipidosis (PL) early in drug discovery and in chemical probe characterization. Here, we present a versatile high-content live-cell imaging approach, which was used to evaluate a chemogenomic and a lysosomal modulation library. We trained and evaluated several machine learning models using the most comprehensive set of publicly available compounds and interpreted the best model using SHapley Additive exPlanations (SHAP). Analysis of high-quality chemical probes extracted from the Chemical Probes Portal using our algorithm revealed that closely related molecules, such as chemical probes and their matched negative controls can differ in their ability to induce PL, highlighting the importance of identifying PL for robust target validation in chemical biology.

INTRODUCTION

Phospholipidosis (PL) is a pathological condition characterized by an excessive accumulation of phospholipids in lysosomes and a distinct morphological appearance of lamellar bodies.[1] PL is frequently induced by the exposure of cells to exogenous substances, including various pharmacological agents.[2] This drug-induced phospholipidosis (DIPL) can have significant implications for drug safety and side effect profiles, including toxicity in the liver and lungs.[3,4] Currently, over fifty clinically studied drugs that are used for treating a broad spectrum of disorders such as neoplastic, psychiatric, and infectious diseases, are known to induce phospholipidosis (Figure 1A).[2,4]

Importantly, DIPL is also a critical risk factor in experiments that utilize chemical probes in cell culture or *in vivo* as the accumulation of phospholipids can have confounding effects in many biological processes.[5] In particular, when using cell-based phenotypic screens, the mechanism of action of the used small molecules needs to be critically assessed.[6] It has been recently reported that several drugs identified in cellular repurposing screens showed apparent antiviral activity against SARS-CoV-

2 via DIPL mechanisms and not by acting on their primary targets.[7] It is therefore of utmost importance to comprehensively characterize libraries used in functional cellular assays not only regarding their general toxicity[8] but also for subtler cellular alterations, such as phospholipidosis induction, to facilitate more reliable target validation in chemical biology.

The exact molecular mechanism underlying the occurrence of PL has not yet been completely elucidated.[2] There are divergent views on whether it is a target-specific process, or whether PL is predominantly due to non-specific effects caused by the chemical properties of the small molecules. For example, the accumulation of small molecules in the lysosomal membrane can lead to the release of lysosomal hydrolases and sphingolipid activator proteins from the surfaces of intralysosomal luminal vesicles.[2] Several targets have been proposed to mediate DIPL, such as lysosome-specific phospholipase A2 (PLA2G15)[9] or the sphingomyelinase (Asm)-ceramide system, but currently no single target can explain all of the observed DIPL cases.[10,11]

DIPL has been associated with cationic amphiphilic drugs (CADs) and the underlying non-specific mechanisms that contribute to phospholipidosis are widely believed to be

Appendix I

Appendix I

Please cite this article in press as: Hu et al., A machine learning and live-cell imaging tool kit uncovers small molecules induced phospholipidosis, Cell Chemical Biology (2023), https://doi.org/10.1016/j.chembiol.2023.09.003

CellPress

Cell Chemical Biology
Article
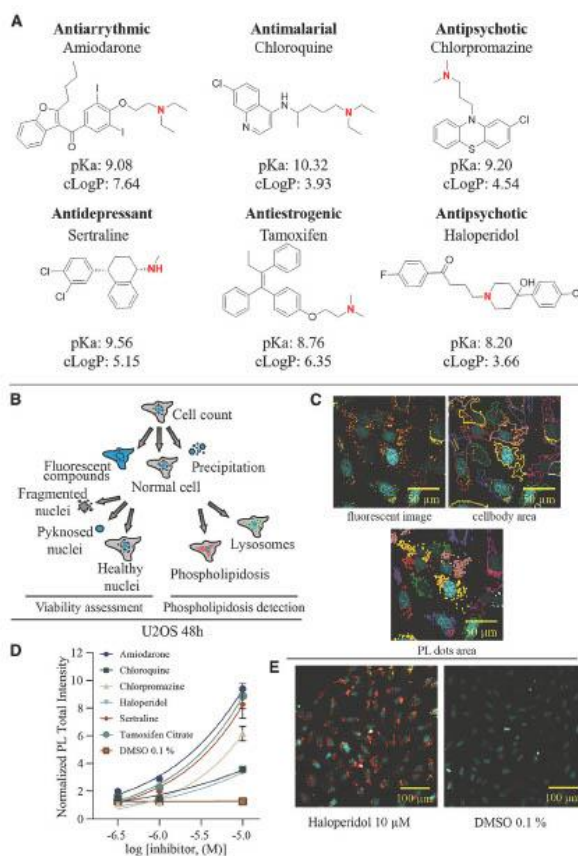
Figure 1. Phospholipidosis detection in cells using compounds known to induce phospholipidosis

(A) Chemical structures with their calculated pKa and cLogP values of drugs known to induce phospholipidosis (amiodarone, chloroquine, chlorpromazine, sertraline, tamoxifen, and haloperidol) and their pharmaceutical application area. Cationic amine group is highlighted in red.

(B) Schematic representation of live-cell phospholipidosis protocol.

(C) Fluorescent images of stained (blue: DNA/Nuclei, green: lysosomes, red: phospholipidosis dots) U2OS cells and the detection of the cellbody area and phospholipidosis (PL) dots after Cellpathfinder analysis.

(D) Normalized PL total intensity of U2OS cells exposed to 0.5 μM, 1 μM, and 10 μM of reference compounds (amiodarone, chloroquine, chlorpromazine, sertraline, tamoxifen, and haloperidol) in comparison to cells exposed to 0.1% DMSO after 48 h. Error bars show standard error of mean (SEM) of four biological replicates.

(E) Representative fluorescent images of stained (blue: DNA/Nuclei, red: phospholipidosis dots) U2OS cells treated with a PL inducer (haloperidol at 10 μM) in comparison to cells exposed to 0.1% DMSO.

associated with the physicochemical properties of CADs. Specifically, CADs feature hydrophobic and hydrophilic domains and carry a positive charge at physiological pH which may contribute to their accumulation within lysosomes.[2] This physicochemical property could be attributed to the ability of small molecules to diffuse passively across lipid bilayers in their unionized form and subsequently become ionized and trapped within lysosomes due to the lower pH values in this cellular compartment.[12,13] For these cases, it is also evident that the binding of CADs to phospholipids may lead to the formation of drug-lipid complexes that cannot be digested.[14,15] Given the impact of PL for drug development, several approaches have been developed to identify molecules that induce phospholipidosis in order to deprioritize them during the early stages of drug discovery.[16] Electron microscopy (EM) has been recognized as the most reli-

able experimental technique for detecting PL.[17] Despite its effectiveness, the widespread utilization of EM has been limited due to the low throughput, high cost, and prolonged timescales of this analysis. Consequently, high-content imaging techniques have gained popularity as a viable alternative. Among the most commonly employed detection techniques are fluorescently labeled phospholipids such as (1-acyl-2-{12[(7-nitro-2-1-,3-benzoxadiazol-4-yl)amino]dodecanoyl} (NBD) NBD-PE staining[18] or Nile red lipid stain.[19] In addition, the fluorescence microscopy-based lysosomal perturbation assay that employs a red fluorescent probe, developed by Coleman et al. in 2010,[20] has also gained widespread adoption as a reliable detection method. However, the implementation of an efficient assay using live-cell imaging methods that could also assess kinetic aspects is still lacking.

Experimental phospholipidosis profiling can be resource intensive, and thus several in silico approaches have been developed to reduce the experimental burden.[16,21] Ploemen et al. introduced the first physicochemical property model that utilized cLogP and pKa as physicochemical properties to identify cationic amphiphilic compounds inducing PL.[22] This method has been widely adopted, and it has been later refined incorporating other properties such as the net charge[23] or volume of distribution (Vd).[24] In general, physicochemical models are useful when applied to small and homogeneous datasets but their predictive capacity is considerably diminished with assessing larger datasets.[25] To further enhance the predictability of phospholipidosis, Przybylak et al. introduced the

2   Cell Chemical Biology 30, 1–18, December 21, 2023

254

Cell Chemical Biology
Article

CellPress

structural alert model comprising 39 structural fragments that closely captured chemical features associated with phospholipidosis, such as peripheral amine groups and hydrophobic moieties.[26,27] Next, scientists from the U.S.A. National Institute of Health (NIH) developed a machine learning (ML) approach based on support vector machine (SVM) model, trained using molecular descriptors of atom types and whole molecule properties of three large compound collections,[28] achieving a high accuracy of 0.90 evaluated by the area under the receiver operator characteristic curve (AUC). NIH scientists also showed that many phospholipidosis inducers and hERG-blocking compounds share a similar structural feature composed by two aromatic rings and an amino group forming a three-center toxicophore.[29,30]

More recently, researchers at AstraZeneca utilized a significant amount of proprietary data (1,537 compounds with 441 inducers) to create a consensus computational model that combines ML and physicochemical property model (Ploemen's criterion).[31] The established model demonstrated strong performance in an external set of 183 compounds (identified 75% of all PL inducers). However, most machine learning approaches studies do not yet benefit from recently developed explainable approaches that allow for model interpretability to better understand their strengths and limitations.[32] Thus, it is important to enhance existing computational approaches and gain deeper insights into the molecular mechanisms underlying phospholipidosis.

In this study, we developed and validated a high-content live-cell imaging approach and screened two diverse libraries comprising ~300 compounds. We also trained a set of ML models using the most comprehensive set of publicly available compounds to date and assessed these models using the two internal experimentally screened sets. We interpreted the best model using the SHAP approach and used it to predict whether high-quality selective probes recommended by the Chemical Probes Portal (https://www.chemicalprobes.org/)[33] may induce PL. Experimental validation confirmed our predictions and allowed us to identify chemical probes that induce PL at recommended concentrations. Finally, our analysis of structurally highly similar compounds, such as chemical probes and their inactive analogs used as negative controls, demonstrated that small changes in compound structure may lead to varying PL annotations. Our findings have important implications for the practice of chemical biology evaluation and characterization, advancing our understanding of the complex nature of phospholipidosis induction.

RESULTS

High-content assay establishment and screening two in-house datasets

Several high-content assays have been proposed in the literature to detect phospholipidosis in a cellular context.[1,18,19] However, these assays are endpoint assays, and they have not been adapted for imaging in live cells. To this end, we used the LipidTox red reagent, an already implemented high-content imaging method first described by Nioi et al.[34] and later optimized by Shahane et al.[1] In this method, phospholipids are conjugated with a fluorescent dye to increase their visibility when they accumulate during phospholipidosis. In order to adapt this assay for live-cell monitoring, we incorporated this detection method into our previously published live-cell multiplex assay and further

optimized this procedure (Figure 1B) (see STAR methods for details).[8] The categorization of PL dots encompassed their localization within both the cell body and lysosomes, with the latter being identified using LysoTracker as an additional dye (Figure 1C). To rule out possible toxicity of the dyes used in live cells over long periods of time (up to 72 h) and to determine the optimal concentration, we performed a viability assessment using alamarBlue after adding different dye concentrations. This experiment revealed no significant effects on cell survival at the concentrations used (Figure S1A). To validate the assay, we tested six drugs known to induce phospholipidosis (Figure 1A)[2-7] including antidepressant drug sertraline, a selective serotonin reuptake inhibitor, antipsychotic drug haloperidol targeting dopamine D_2 receptor, and the antiestrogenic agent tamoxifen using osteosarcoma (U2OS) cells. Drugs were tested at three different concentrations over 72 h, and the normalized PL was assessed based on the total intensity of the dyed molecules compared to cells treated with DMSO. After 48 h, all reference compounds significantly induced PL, which was detected already at the lowest concentration used (500 nM) (Figure 1D). For all control compounds, induction of PL correlated with compound concentration and the time of exposure (Figure 1E). To enable subsequent machine learning analysis (a binary classification task mirroring the public compound annotation used for model training) of in-house developed datasets, it was necessary to establish a rigorous threshold for classifying the compounds as either "inducer" or "non-inducer" of PL. Utilizing the initial reference dataset, we established a threshold in U2OS cells by defining normalized PL intensity larger than 2 based on the intensity measured for chloroquine (Figure 1D). Nonetheless, in cases where compounds were close to the threshold value, we performed a careful visual assessment to determine whether the compound should be categorized as an "inducer" or "non-inducer". Most PL data described in the literature have been recorded in HepG2 cells. Since our assay was performed in osteosarcoma cells (U2OS), we tested the context dependency of PL in different cell lines by testing the reference compounds also in HEK293T and HepG2 cells after 24 h (Table S1). In all cell lines, PL was detected for all reference compounds. Given the distinct morphological characteristics of each cell line, a direct comparison of the total intensities was not feasible (Figure S1B). We therefore recommend considering each cell line separately and setting the threshold for significant PL induction accordingly. This will be particularly important in drug discovery projects where PL may be evaluated on patient-derived cell models to assess the corresponding intensity threshold appropriately.

After establishing the high-content assay and defining the threshold to distinguish between "inducer" and "non-inducer", we tested two in-house compound sets in order to identify PL inducers. The first compound set included a diverse set of chemogenomic compounds targeting proteins of different families, including GPCRs, kinases, epigenetic modulators, and others (Table S2A). In contrast to chemical probes which are highly selective and potent modulators of their targets, frequently accompanied by matched structurally related negative controls, chemogenomic compounds are not exclusively selective and they often target more than one protein. However, the selectivity of chemogenomic compounds has been comprehensively assessed, and the hope is that phenotypic changes in cell culture
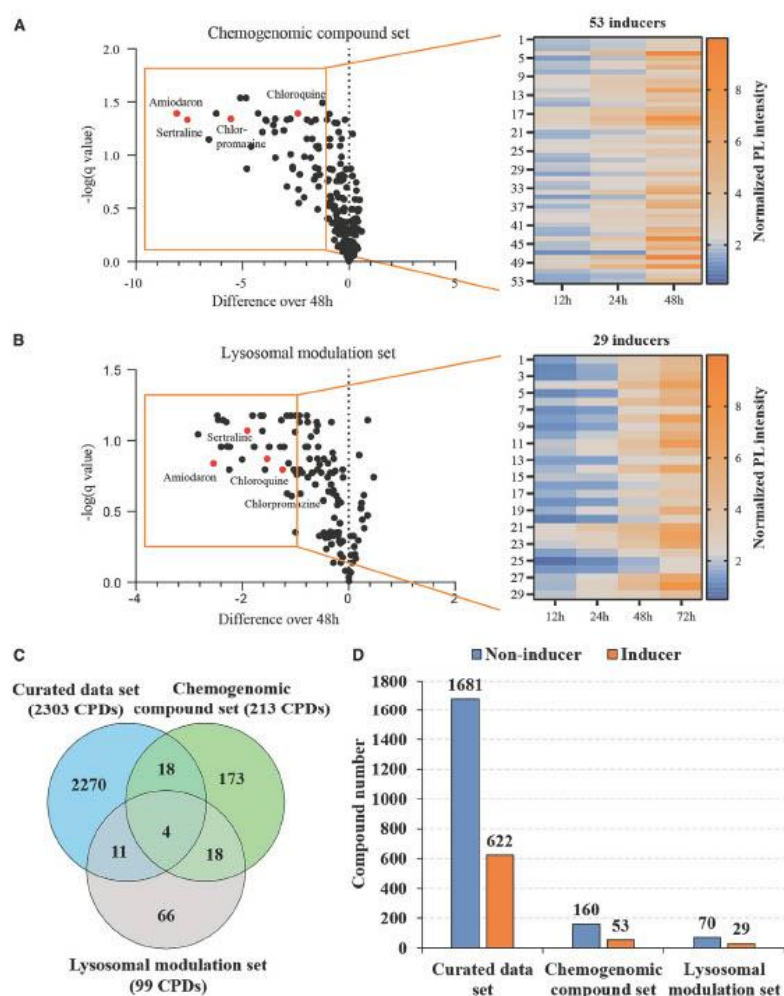
Figure 2. Live-cell assay to distinguish phospholipidosis inducer and non-inducers and the screening of two in-house datasets
(A) Volcano plot of compounds in the chemogenomic compound set tested for phospholipidosis inducers or non-inducers in a live-cell assay using U2OS cells over 48 h. Each dot represents a compound tested at 10 μM. The x axis shows the log 2-fold change in normalized PL total intensity between cells treated over 6 h and cells treated over 48 h. The y axis shows the negative log (q value) of the adjusted p value for each compound. Reference compounds are shown in red. The orange square indicates compounds that are defined as "inducers" for significant phospholipidosis induction (normalized PL intensity >2). 53 "inducers" compounds are highlighted in a heatmap showing normalized PL intensity after 12 h, 24 h and 48 h of compound treatment, with orange representing high induction and blue representing low induction. Data were normalized against the average of DMSO (0.1%)-treated cells. All experiments were performed in biological duplicates. Significance was calculated using an unpaired multiple test analysis using GraphPad Prism 8.4.3.

(legend continued on next page)

Cell Chemical Biology
Article

CellPress

can be assigned to a target based on hits from chemogenomic compounds with similar on-target activity but diverse off-target profiles. The chemogenomic compound set consisted of 213 compounds, out of which 53 (25%) were identified as "inducers" with a normalized PL intensity larger than 2 (Figure 2A; Table S2B). To assess compounds, that are known to play a pivotal role in lysosomal modulation, a second set consisting of 99 compounds was used (Table S3A). In order to capture cellular phenotypes over long incubation times, this set was subjected to a longer incubation time of 72 h. In total, 29 compounds (29%) of this second set were defined as "inducers" (Figure 2B, Table S3B). Among the PL-inducing compounds was crizotinib, an approved receptor tyrosine kinase inhibitor targeting ALK and cMET, as well as a number of other protein kinases, which have been reported to boost the lysosomal count.[35] In addition, we detected PL upon exposure of cells with daurisoline, which acts as an autophagy inhibitor suppressing lysosomal acidification.[36] Also, the histone deacetylase (HDAC) inhibitor mocetinostat, which previously has been shown to trigger accumulation of lysosomes in cells induced PL.[37] As an external control, four reference compounds (amiodarone, chloroquine, chlorpromazine, and sertraline) were added to both sets and were correctly identified as "inducers", demonstrating the reproducibility of our live-cell assay. There were 22 compounds that were present in both sets, which acted as internal controls (Figure 2C). Interestingly, although for most compounds phospholipidosis was observed after 24 h in both sets, some compounds reached the threshold only at 48 h, indicating the importance of assessing compounds at longer time points.

PL has been frequently associated with off-target phenotypic effects, and it might contribute to toxicity of diverse approved drugs.[5] Nevertheless, in our 2-D cell culture models, we did not observe correlation between PL and cytotoxicity, and only few compounds, such as kinase inhibitor adavosertib, showed experimentally PL induction at late time points and affected cell viability (Figures S2A and S2B). Other authors[4] have suggested a connection between autophagy and PL. However, following the assessment of the lysosomal modulation set in an LC3 HiBiT reporter assay, we did not observe a direct correlation between phospholipidosis and autophagy (Table S4). Therefore, additional evaluations may be required to further elucidate the relationship between these two phenomena.

Curation of a comprehensive phospholipidosis dataset from the literature

First, a thorough analysis of PubMed was carried out, resulting in the identification of nine articles that reported large compound datasets assessed for PL. From these articles, we identified reported compounds, downloaded their chemical structure with the aid of the PubChemPy package (https://github.com/mcs07/PubChemPy) and applied a structural standardization protocol (see STAR methods for details). This analysis resulted in a set of 2,422 unique compounds that we termed the curated dataset. Among these compounds, 114 (4.7%) had inconsistent phospholipidosis annotations in the literature, a fact that has been observed previously.[28] For example, a number of compounds have been labeled as inducers as well as non-inducers in different publications. These inconsistencies illustrate the challenges around comprehensively identifying phospholipidosis and building computational models with the limited data quality, quantity, and availability. To attain data with a high level of confidence, we eliminated compounds that possessed inconsistent annotations. This led to the creation of an initial set consisting of 2,308 literature compounds, of which 1,683 were non-inducers and 625 were inducers.

Next, we assessed the overlap between the curated set and our two in-house compound sets. This analysis identified 38 compounds that were present in both sets. Of these, only five compounds (~13%) had inconsistent annotations due to precipitation or high toxicity and were excluded from the analysis. Overall, a total of 213 compounds for the chemogenomic compound set (Table S2), 99 compounds for the lysosomal modulation set (Table S3), and 2,303 compounds for the curated literature set (Table S5) were obtained. The Venn diagram in Figure 2C presents an overview of the distribution of compounds among these three sets, while Figure 2D illustrates the distribution of phospholipidosis compositions in each set. Specifically, the curated set obtained from the literature contained 622 inducers and 1,681 non-inducers; the chemogenomic compound set comprised 53 inducers and 160 non-inducers; and the lysosomal modulation set consisted of 29 inducers and 70 non-inducers. In general, the percentage of inducers in each set was quite similar with 24.9%–29.3% PL inducers.

A simple physicochemical model is limited to differentiate inducers from non-inducers

Using the curated dataset obtained from the literature, we initially evaluated the effectiveness of a simple physicochemical property model (Ploemen's criteria using pKa and cLogP values)[22] to differentiate phospholipidosis inducers from non-inducers (see STAR methods for details). An overview of the analytical workflow is presented in Figure 3. Using this simple model, an accuracy value of 0.74 was obtained for the curated set of 2,303 compounds meaning that 74% of PL inducers and non-inducers were correctly predicted. However, we only obtained a recall value of 0.35 (Figure 3, left), indicating that this simple model was only able to identify 35% of all PL inducers and illustrating the model's limited ability to comprehensively predict phospholipidosis (i.e., inducers were frequently misclassified as non-inducers). The suboptimal performance of the model prompted us to explore whether ML algorithms could be more effective.

A random forest algorithm coupled with concatenated features outperforms other approaches

Firstly, we evaluated four distinct ML algorithms, namely k-nearest neighbor (KNN), support vector machine (SVM), random

(B) Volcano plot of compounds in the lysosomal modulation set tested for phospholipidosis inducers or non-inducers in a live-cell assay using U2OS cells over 48 h. 29 "inducers" compounds are highlighted in a heatmap showing normalized PL intensity after 12 h, 24 h, 48 h, and 72 h of compound treatment.

(C) Schematic overview of the different datasets used in this study (curated dataset, chemogenomic compound set, lysosomal modulation set) and the number of overlapping compounds.

(D) Bar diagram of different datasets used in this study (curated dataset, chemogenomic compound set, lysosomal modulation set). Bars with a blue color are showing the number of compounds defined as "non-inducer", while orange bars show the number of "inducers" of every compound set.

CellPress

Cell Chemical Biology
Article
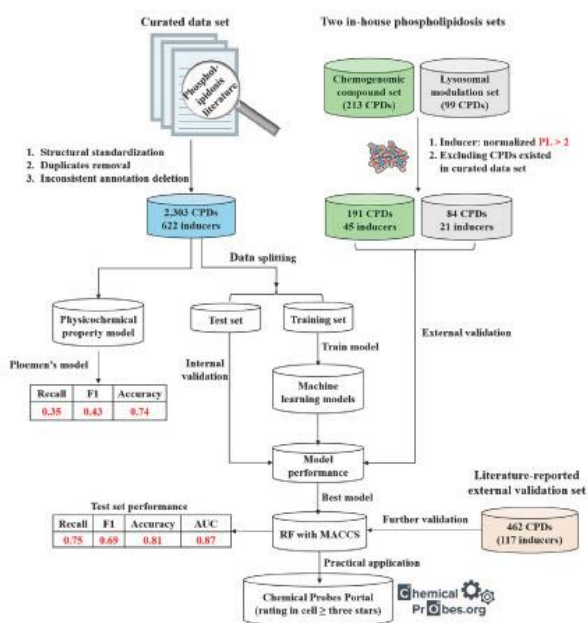
Figure 3. Flowchart of the analysis

The process began with a curation of reported phospholipidosis data from scientific literature, followed by the application of a multistep structural standardization protocol to obtain high-confidence data for model construction. Initially, a simple physicochemical property model was applied to the curated dataset. However, it was found to have unsatisfactory performance, indicating the need for a more robust model capable of accurately distinguishing between inducers and non-inducers. Therefore, ML models were subsequently developed. To validate the performance of the ML model, two in-house screening datasets were utilized—a chemogenomic compound set and a lysosomal modulation set, in addition to a test set. The evaluation metrics were used to select the best ML model which was further validated by additional PL data from the literature. The best model was subsequently used for practical applications in predicting chemical probes in the Chemical Probes Portal. The metric of recall assesses the model's ability to correctly identify positive samples, while the F1 score is a measure of the harmonic mean of precision and recall. Accuracy provides an overall estimate of the model's prediction accuracy.

forest (RF), and eXtreme gradient oosting (XGBoost), using MACCS and ECFP4 fingerprints as molecular representations. In order to better recapitulate the physicochemical properties that are known to be important to predict PL, the molecular representations were concatenated with two additional properties (pKa and cLogP) and termed "concatenated features" (see STAR methods for details). At first, the curated data were split into two distinct datasets: a training set that served the purpose of training the model and fine-tuning its hyperparameters and a testing set that was utilized to evaluate the model's performance internally. Moreover, the two physical phospholipidosis sets evaluated in-house were used as external validation to assess the model's predictability on real-world datasets, as depicted in the workflow illustrated in Figure 3. The results presented in Figures 4A–4C and S3 indicate the performance of the ML models across 10 independent trials (i.e., 10 times independently splitting the curated data to train and test the models). Our analysis revealed that most of the models had robust predictive performance on the test set, with the RF model using MACCS fingerprint exhibiting the highest average recall value of 0.73. Accuracy values around 0.80 and F1 values around 0.6 were similar across all explored ML methods (Figure 4A). We next investigated the generalization ability of these models by assessing their performance on the chemogenomic compound set and the lysosomal modulation set (external validation datasets). Notably, the RF algorithm coupled with MACCS fingerprint

also outperformed the SVM, KNN, and XGBoost models in this setting, as indicated by the highest average recall and F1 scores of 0.57 and 0.47, respectively, on the chemogenomic compound set (Figure 4B). In contrast to the RF algorithm, SVM, KNN, and XGBoost exhibited average recall values of less than 0.4 with any of the two fingerprints tested. Similarly, when evaluating the lysosomal modulation set (Figure 4C), the random forest algorithm using MACCS fingerprint exhibited superior performance over the other three methods, achieving an average recall value of 0.84.

Following our assessment of the three sets, we decided to further investigate the performance of the RF algorithm in combination with different molecular representations, including MOE 2D descriptors (206 physicochemical property features) and PubChem fingerprint (881 chemical substructure patterns). Our analysis demonstrated that there was little variation in the predictive performance of the RF algorithm when using different molecular representations in both the test set (Figure S3A) and our two in-house sets (Figures S3B and S3C).

Based on these results and also considering computational efficiency, we decided to focus on RF in combination with the MACCS fingerprint for further analysis. To further assess this model, we plotted the ROC curves of the three investigated sets based on this best ML model (Figure 4D). The AUC values of the test set and lysosomal modulation set were 0.87 and 0.90, respectively—a high accuracy comparable to previously reported methods.[28] These values suggested that the model has a high-predictive capability for these two sets. However, the models' predictability for the diverse chemogenomic compound set was lower with an AUC value of approximately 0.7, indicating that it is less able to make accurate predictions for this set.

Cell Chemical Biology
Article

CellPress
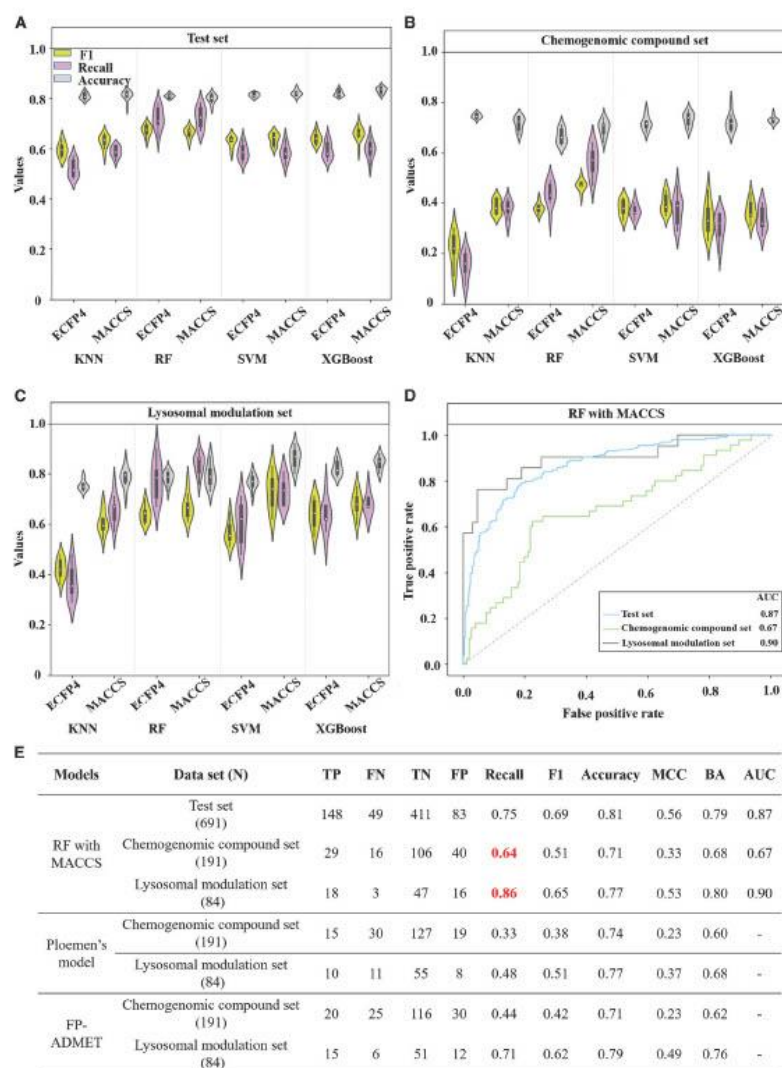
Figure 4. Performance machine leaning models in combination with different molecular representations
(A) The violin plots illustrate the performance of four different ML algorithms, which were based on MACCS and ECFP4 fingerprints concatenated with calculated LogP and pKa. The models were evaluated using the test set across 10 independent trials.
(B) The models were evaluated using the chemogenomic compound set.
(C) The models were evaluated using the lysosomal modulation set.

The statistical results of the RF model with MACCS on the test set, chemogenomic compound, and lysosomal modulation sets based on the best trial are presented in Figure 4E. For comparison, the same sets were predicted using Ploemen's simple physicochemical property model. As shown in Figure 4E, both the physicochemical and RF models have similar accuracy values (0.71–0.81) while the F1 and, importantly, the recall, are much better with the RF (0.64–0.86) than with the physicochemical model (0.33–0.48). These results demonstrate that the developed ML model can identify many more PL compounds that would be misclassified if only physicochemical properties were considered.

To benchmark our ML model further, we applied previously published methods that reported the exact code or were straightforward to use to predict new datasets (e.g., are available via webserver) and we identified FP-ADMET,[38] a repository of molecular fingerprint-based RF built comprehensively exploring molecular descriptors for several ADMET properties, including PL prediction (see STAR methods for details). FP-ADMET built an RF model to predict PL using 1,719 compounds and reported an AUC of 0.88. When we used FP-ADMET to predict our chemogenomic and lysosomal modulation sets, its recall was 0.44 and 0.71, respectively (Figure 4E). It is worth highlighting that although FP-ADMET was able to improve the recall of the simple Ploemen's physicochemical property model (0.33, 0.48), our algorithm displayed a significantly better recall (0.64, 0.86) on these two external validation sets—illustrating how our algorithm compares favorably with available methods and can identify more DIPL compounds from two different compound sets.

As a means of an additional validation, we identified a set of 462 new compounds (117 PL inducers and 345 non-inducers, Table S5) from public sources,[1,29] which we used as an additional validation set. Using our RF model, we were able to correctly predict 85 of the compounds as PL inducers, which correspond to a recall value of ~0.73—in line with our previous recall values of 0.64–0.86 in the external validation sets (Figure 4). The remaining performance metrics were also in line with the results on the two previous external validation sets or slightly higher (Figure S4); hence further highlighting the strong predictability and robustness of our algorithm when validated prospectively and when we compared it to other published approaches such as FP-ADMET.

Two different real-world prediction scenarios explain the disparity of prediction performances

Intrigued by the marked discrepancy in the predictive ability of the ML method developed when applied to the chemogenomic compound (recall = 0.64) or lysosomal modulation (recall = 0.86) sets (Figure 4E), we decided to explore the underlying factors.

First, we evaluated the chemical similarity between the training set and the three investigated sets (Figure S5). As expected, the training set (used for model construction) had the highest chemical similarity to the test set with a median maximum Tanimoto coefficient (mmTc) of 0.51, followed by the lysosomal modulation set (mmTC = 0.33). The chemogenomic set was the most dissimilar (mmTc = 0.29), which may explain why all methods perform worse with this external dataset (Figure 4E). However, it is worth highlighting that our model still yielded the best performance across all models tested when predicting the chemogenomic set, illustrating its effectiveness in a challenging, real-world scenario.

In addition to the chemical similarity, we also analyzed the physicochemical properties on each set of compounds to gain more insights into the differences between them. As shown in Figure S6, these two sets represent different areas of the property space which translate into different levels of difficulty for real-world computational prediction. In the lysosomal modulation set, most PL inducers fall within the cationic amphiphilic area of the chemical space (cLogP $\geq$ 3 and pKa $\geq$ 7.4). Therefore, they are easy to predict, which explains the higher recall of the ML method (0.86) and that even the simple physicochemical model achieves a significant recall (0.48). In contrast, in the chemogenomic set, there are far more PL-inducing compounds (orange dots in Figure S6) that fall outside the cationic amphiphilic area of the property space and thus are more difficult to predict. Therefore, different external validation sets can be easier or harder to predict depending on the area of the chemical space they cover. Despite these limitations, ML still outperforms the simple model by a substantial margin (Figure 4E). Additionally, we also explored whether incorporating pKa and cLogP properties into the MACCS fingerprint was truly improving the prediction performance. This was clearly demonstrated by the increase in recall values from 0.53 (without pKa and cLogP properties) to 0.64 (with these two properties) on the chemogenomic compound set, as presented in Table S6. Therefore, we made the decision to continue utilizing the RF model with MACCS and concatenated features for further analysis.

Model interpretation based on SHAP-based feature analysis

Based on the selected computational model, it was important to elucidate the underlying mechanisms by which the model generates its predictions to understand its strengths and limitations. Therefore, we applied the SHAP approach[39,40] to analyze the contributions of individual features (see STAR methods for details). The top 15 features that significantly affected the prediction output, along with their SHAP value distributions, were analyzed and presented in Figure 5A. These top features were also mapped back to an exemplary chemical structure to facilitate interpretation, as shown in Figure 5B. Due to the relatively low-SHAP values observed for all feature bits (< 0.1, Figure 5A), it was inferred that there were no dominant features that strongly

(D) Based on the best trial of the RF with MACCS concatenated features, ROC curves for three validation sets are plotted.

(E) The statistical performance measures of the computational models including the simple physicochemical model and FP-ADMET model are presented. TP refers to "true positive", TN to "true negative", FP to "false positive", and FN to "false negative". The metric of recall assesses the model's ability to correctly identify positive samples, while the F1 score is a measure of the harmonic mean of precision and recall. Accuracy provides an overall estimate of the model's prediction accuracy; BA is the arithmetic mean of sensitivity and specificity when dealing with imbalanced data; MCC a metric used to assess the performance of binary classification models, a score of 0 indicates that the classifier performs no better than random guessing. The receiver operating characteristic (ROC) curve and its corresponding area under ROC (AUC) were utilized as a threshold-independent metric to further assess the overall predictive performance. With the exception of the MCC, which ranges from −1 to 1, all other metrics are scaled between 0 and 1, with a score of 1 indicating optimal classification performance.

Cell Chemical Biology
Article

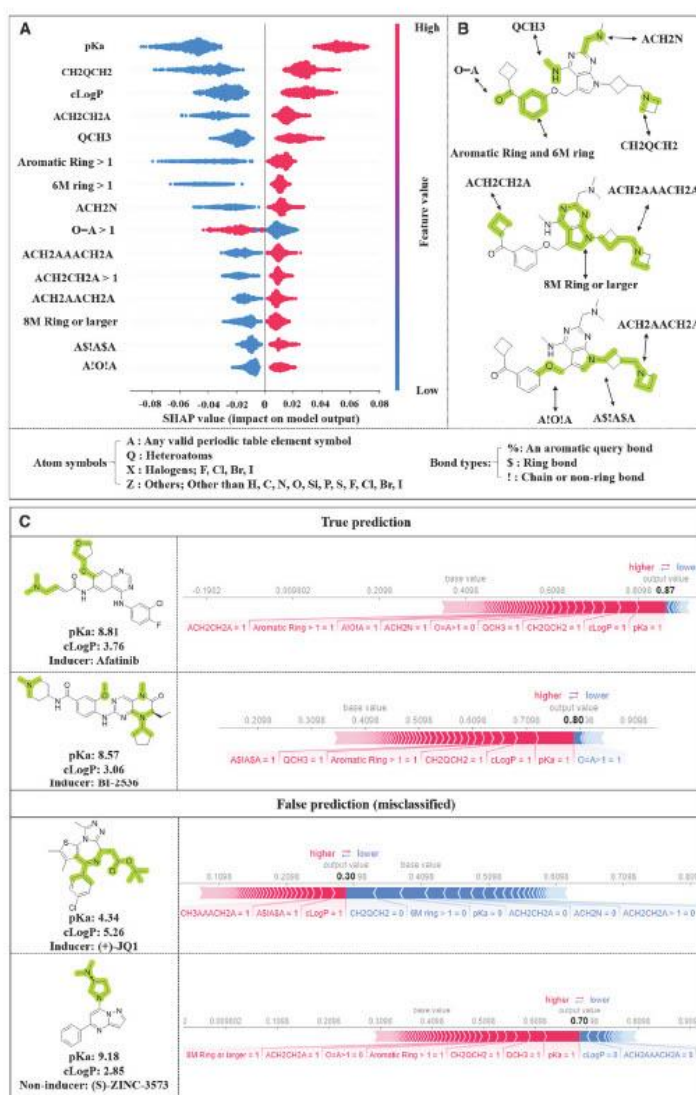**Figure 5. Top ranked features analysis and rationalization of the prediction on exemplary cases**
(A) Shown is the distribution of the top 15 features identified by SHAP analysis, which was conducted using the RF algorithm in conjunction with MACCS (concatenated with cLogP and pKa) for phospholipidosis inducer prediction. Each feature's line is represented by a series of dots, where each dot

(legend continued on next page)

CellPress

Cell Chemical Biology
Article

influenced the predictions generated by the ML model. Thus, using a small number of features may not yield accurate predictions. All top features, except for the MACCS feature "O = A >1", had a positive effect on predicting PL inducers. Thus, the existence of these features additively increases the likelihood of molecules being categorized as inducers. The most influential features were the molecular physicochemical properties of pKa and cLogP, as well as the MACCS fingerprint "CH2QCH2". The presence of the pKa feature (pKa $\geq$ 7.4) and cLogP feature (cLogP $\geq$ 3.0) indicated that these compounds can be protonated to carry a positive charge in the acidic endolysosomal compartment and are more amphiphilic.[41] This observation was in line with previous reports that suggested that compounds that induce phospholipidosis tend to be cationic amphiphilic in nature.[2] Nevertheless, a significant number of compounds inducing PL fall outside the cationic amphiphilic chemical space (Figure S6). The "CH2QCH2" substructure often accounts for the presence of a hydrophilic amine group in a molecule, which is a structural representation of the pKa property. Additionally, other MACCS structural features, such as aromatic or six-membered ring systems, indirectly accounted for the lipophilic nature of molecules. In summary, the ML model outperformed the simple physicochemical model by giving small weights to multiple features that included pKa and cLogP (indicating the characteristics of the whole molecule) as some of the top-ranking features, but the model also ranked highly chemical features representing ionizable and aromatic/aliphatic moieties. Therefore, the ML method "learned" that even if pKa is not $\geq$ 7.4 or cLogP $\geq$ 3, the presence of other chemical moieties contributed to correct predictions of PL.

Next, we attempted to deeply explore the prediction of a few representatives from our in-house compound sets shown in Figure 5C. Afatinib and BI-2536 are kinase inhibitors that target EGFR/ErbB and PLK1, respectively. While Afatinib induces autophagy and thereby increases lysosomal activity,[42] BI-2536 suppresses lysosomal fusion.[43] Nonetheless, both compounds were correctly categorized as inducers by the ML model, with predicted probability values of 0.87 and 0.80, respectively (Figure 5C). The high probability of an inducer prediction can be attributed to the prevalence of several positive features represented by positive SHAP values which are highlighted in lime on their corresponding chemical structures. Our analysis also included two misclassified chemogenomic compounds, namely (+)-JQ1 and (S)-ZINC-3573. (+)-JQ1 is a BET bromodomain inhibitor[44] that suppresses cell proliferation and promotes lysosomal modulation. The model wrongly predicts that (+)-JQ1 is not a phospholipidosis inducer, which is primarily attributed to the absence of significant features, such as a hydrophilic amine

group with a pKa value below 7.4 (Figure 5C). We also analyzed the features of the 18 false negative (FN) cases (Figure 4E and Table S7) in the chemogenomic compound and lysosomal modulation sets. The majority of them (89%) did not contain the "pKa $\geq$ 7.4" feature, indicating available moieties were not predicted to be sufficiently protonated. These results could prompt the development of better pKa prediction approaches or challenge the necessity of compounds to be protonated in order to induce PL. On the other hand, a certain number of false positive (FP) cases were also identified, such as (S)-ZINC-3573. (S)-ZINC-3573 is the inactive enantiomer of chemical probe (R)-ZINC-3573, a selective agonist of MRGPRX2.[45] (S)-ZINC-3573 was misclassified as an inducer of phospholipidosis because it presents several features such as aromatic rings, "CH2QCH2", and pKa $\geq$ 7.4. These two examples illustrated that the presence or absence of specific moieties and physicochemical properties was not sufficient to correctly classify all compounds. Taken together, the SHAP values assigned to each feature contribute to the transparency of the ML prediction, significantly aiding our comprehension of how the model arrived at the specific prediction.

Structurally similar compounds may differ in inducing phospholipidosis

Intrigued by the existence of compounds that produced a phospholipidosis phenotype that was contradictory with the general pKa, cLogP, and features associated with it, we decided to investigate the structural relationship between compounds with varying phospholipidosis annotations through analog analysis. In this study, we used a substructure-based approach known as the matched molecular pair (MMP)[46] to identify structural analogs among the investigated three compound sets. MMPs can be obtained via the application of a compound fragmentation algorithm (see STAR methods for details).[47] Through a systematic fragmentation of compounds, a total of 1,368 MMPs were generated, comprising 1,149 unique compounds (1,078 of which were chemogenomic/lysosomal modulation compounds). This corresponds to approximately 45% of the entire collection of compounds analyzed indicating high-structural diversity in the three sets. To visualize the generated MMPs, they were organized into a network (Figure 6, upper panel). In this network, the nodes correspond to compounds and the edges represent the pairwise MMP relationship. Coordinated MMPs formed by subsets of compounds gave rise to the formation of distinct clusters within the network. Upon analyzing the generated MMPs, we were able to identify a total of 343 clusters. In general, cluster sizes are small, with an average of three to four compounds per cluster. Out of the identified clusters, a

corresponds to a molecule, and the stacking of dots shows the density of SHAP values. A higher positive SHAP value suggests a more significant contribution to the prediction of phospholipidosis inducers. Atom symbols and bond types are provided at the bottom of the figure for reference.
(B) To aid in understanding the MACCS substructure patterns, the top 13 features (excluding pKa and cLogP) have been mapped back to a chemical structure, which is highlighted in lime. For clarity, some MACCS features such as "ACH2CH2A" specifying an atomic environment via using a recursive SMARTS pattern (https://github.com/rdkit/rdkit/blob/master/rdkit/Chem/MACCSkeys.py), the exemplary structural environment is also highlighted alongside the atom.
(C) Four exemplary compounds coming from in-house datasets depict how the model predicts the induction of phospholipidosis. Two of these compounds (top) were correctly predicted, while the other two (bottom) were misclassified. The contribution of each feature is presented alongside the SHAP force plot, which highlights the features that positively or negatively contribute to the inducer prediction with red and blue colors, respectively. In the force plot, features with a value of zero or one indicating the absence or presence of the corresponding feature and the length of the bars indicates the degree of contribution of each feature toward the prediction. The base value is the base probability and the number with black bold font in plot represents the final output (predicted probability to be PL inducer) of all features.

CellPress
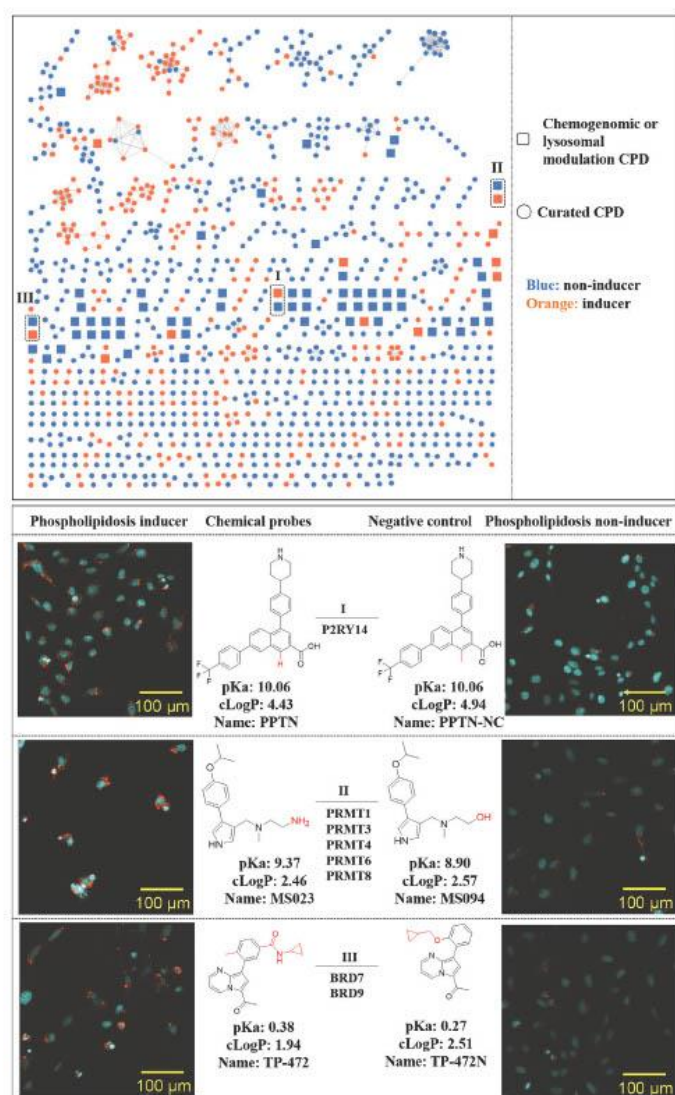
Figure 6. Structure-phospholipidosis network and exemplary analog pairs with disparate phospholipidosis annotation
The network representation at the top displays MMP-based relationships between compounds from three sets. Each node represents a compound and the edges represent pairwise MMP relationships. The circular and rectangular nodes correspond to compounds from curated and in-house sets, respectively. The

263

significant portion of 213 (62.1%) consisted solely of non-inducers while a smaller subset of 43 (12.5%) was composed exclusively of inducers. This suggests that compounds with similar structural characteristics tend to exhibit similar phospholipidosis profiles. However, we also detected 87 clusters (25.4%) that contained compounds exhibiting both phospholipidosis inducer and non-inducer characteristics indicating that structurally similar compounds may significantly differ in their phospholipidosis induction properties. The analysis of analogs with varying phospholipidosis annotations yielded a total of 236 MMPs (17.2%). Among the MMPs, we have identified a total of 214 distinct chemical transformations (Table S8), representing variations in their chemical composition. The large number of transformations suggests that there is not a single or a few structural modifications that are able to unambiguously explain PL induction. The most frequent transformation is an H ↔ CH3 transformation (7 MMPs), followed by an H ↔ OH transformation (5 MMPs) and an H ↔ F transformation (4 MMPs). All of them are single-atom modification indicating that minor structural modifications with conservative impact on physicochemical properties are enough to alter the PL behaviors of small molecules. This observation underscores the intricate nature of the PL mechanisms at play. These aforementioned results indicated that protonation state, pKa, cLogP, and substructural features were insufficient to correctly identify all PL-inducing compounds, which has important implications for the development of more robust computational methods and highlights the complexity of PL.

Chemical probes and their corresponding negative controls may induce divergent phospholipidosis phenotypes

It is noteworthy that out of the 236 MMPs with different phospholipidosis annotations, 13 MMPs were related to compounds that we screened in-house. Intriguingly, four MMPs were formed by a chemical probe and the corresponding structurally related negative control, while the fifth MMP was formed by two chemical probes (GSK6853 and PFI-4). Three of them were highlighted in the network (cluster I, II, and III) and were summarized in Figure 6.

In three cases, chemical probes were experimentally found to induce phospholipidosis, while their corresponding negative controls were non-inducers. Despite variations in activity against their intended primary targets (controls are inactive or much less active), pairs consisting of chemical probes and negative controls frequently displayed similar physicochemical properties, including comparable pKa and cLogP values. Therefore, very small chemical modifications can transform a compound from a PL inducer to a non-inducer without significantly affecting physicochemical properties. For example, the chemical probe PPTN functions as an antagonist of the GPCR family member P2RY14 (IC_{50} = 0.5 nM).[48] The mere substitution of a hydrogen atom with a methyl group nearly abrogates affinity for P2RY14 (IC_{50} = 4,381 nM, >

8,000-fold difference) while pKa and cLogP remain very similar. MS023 (targeting PRMT1, PRMT3, PRMT4, PRMT6, and PRMT8)[49] and TP-472 (targeting BRD7 and BRD9)[50] represent two additional examples of the same behavior.

Another interesting chemical probe/negative control pair was TP-064/TP-064N (Figure S7) targeting PRMT4.[51] In this case, despite replacing the methoxy group with a secondary amine, both compounds remained inducers—demonstrating that these small structural changes did not necessarily induce a change in the propensity of small molecules inducing PL. Finally, in the case of SGC3027 (Figure S7), which targets PRMT7,[52] we observe the opposite behavior where the probe was experimentally confirmed to be a non-inducer while the negative control was an inducer of PL.

Overall, the identification of this series of chemical probes/negative control pairs revealing differences in phospholipidosis induction, demonstrated that minor chemical modifications with very limited impact on pKa/cLogP values and overall compound structure can transform a compound from a PL inducer to a non-inducer. These small alterations in compound structure pose challenges on computational predictions explaining why overall compound properties (pKa, cLogP) were insufficient in predicting PL in some cases. These results also caution on the use of negative controls without further validation or characterization since the observed phenotype differences between probes and controls can be mediated via PL and not the known drug target.

PREDICTION AND EXPERIMENTAL CONFIRMATION OF CHEMICAL PROBES INDUCING PHOSPHOLIPIDOSIS CHALLENGE THEIR USE AT RECOMMENDED HIGH CONCENTRATIONS

Finally, we decided to apply our ML model to identify chemical probes that might be inadvertently producing PL at high concentrations and thus represent a risk when they are used for target validation. To identify high-quality probes, we used the Chemical Probes Portal (https://www.chemicalprobes.org/)[33] and predicted all of them using our ML algorithm. Of these high-quality probes, 160 probes were predicted to be PL inducers. We then prioritized 31 probes with predicted probability value greater than or equal to 0.7 to induce PL for experimental validation (see STAR methods for details). From these, we correctly predicted 21 chemical probes as PL inducers, resulting in a precision value of around 0.7. Table S9 provides a comprehensive list of the 31 probes that were tested and experimental raw data are provided in Table S10. Figure 7A summarizes the primary targets of these 21 chemical probes. In total, 31 targets covering epigenetic and kinase groups were identified illustrating that there is no prominent primary target that is shared or increases the likelihood of compounds inducing PL. Figure 7B depicts 10 of these chemical probes that we predicted and experimentally confirmed to induce phospholipidosis and are

orange and blue hues of the nodes denote compounds that act as phospholipidosis inducers and non-inducers, respectively. At the bottom, three analog pairs consisting of chemical probes and their corresponding negative controls (highlighted in the network) are presented. Each pair is accompanied by the compound name and two calculated physicochemical properties (cLogP and pKa). The primary target gene names corresponding to each probe are provided in the center. Structural modifications between pairs are highlighted in red. Additionally, fluorescent stained image (blue: DNA/Nuclei, red: phospholipidosis dots) of U2OS cells for each compound is also provided.

Cell Chemical Biology
Article

⌀ CellPress
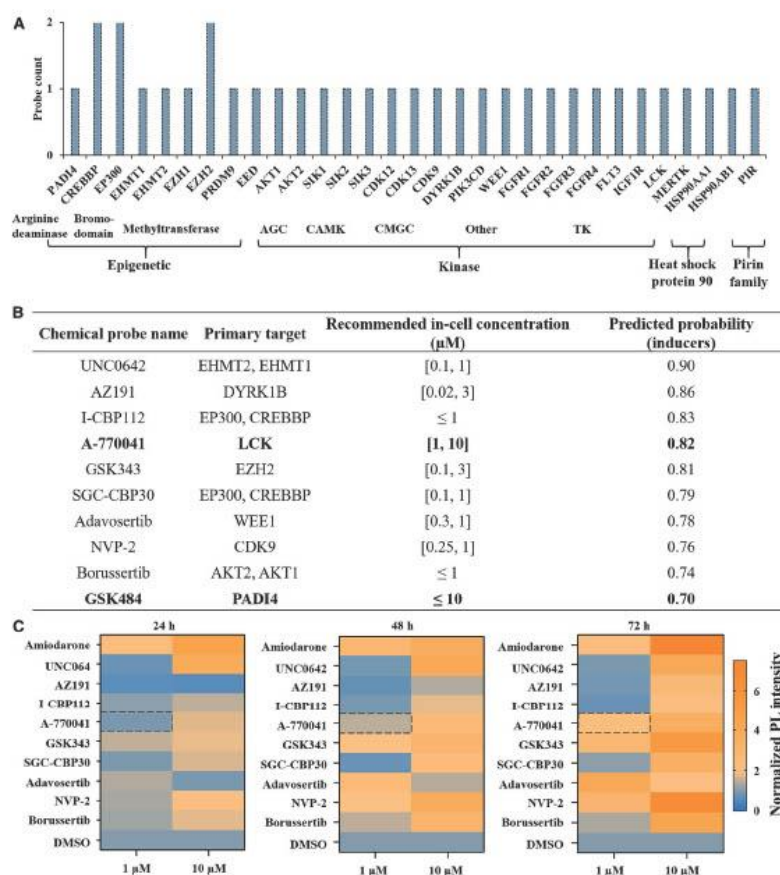
Figure 7. Prediction and experimental validation of high-quality chemical probes from the Chemical Probes Portal inducing phospholipidosis

(A) Target family distribution for the 21 probes confirmed to induce phospholipidosis.

(B) Details of 10 chemical probes correctly predicted by the ML model and experimentally confirmed to be phospholipidosis inducers recommended for use at high concentrations (1 μM or 10 μM in cell-based assays). Each probe is accompanied by its name, primary target (gene name), recommended in-cell concentration by the Chemical Probes Portal, and predicted probability of inducing PL.

(C) Heatmap showing normalized PL intensity in U2OS cells after 24 h, 48 h and 72 h of selected compounds treatment at two concentrations (1 μM and 10 μM) in comparison to cells treated with 0.1 % DMSO. Compound A-770041 is highlighted as a chemical probe that is recommended at higher concentrations (1–10 μM) and that induces PL at longer time points. Orange is representing high-PL induction and blue is representing low-PL induction. Values show SEM of two biological duplicates.

recommended for utilization at high concentrations (up to 10 μM) by the Chemical Probes Portal. The lower panel highlights the time and concentration-dependent progression of phospholipidosis induction, revealing an augmented induction at higher doses and prolonged exposure periods. These observations underscored the necessity of carefully accounting for such dose and time-dependent effects when employing these probes in experimental settings. Alarmingly, A-770041 and GSK484

Cell Chemical Biology 30, 1–18, December 21, 2023   13

265

CellPress

induce phospholipidosis at their recommended 10 µM dose. In addition, A-770041, GSK343, adavosertib, and NVP-2 induced phospholipidosis at even lower concentrations (1 µM) (Figure 7C) that also fall within the portal recommended concentrations. Therefore, when utilizing these probes at the suggested concentrations in functional cell assays, a contribution of phospholipidosis to the observed phenotype should be taken into consideration. We have contacted the Portal with a request to annotate the compounds accordingly.

DISCUSSION

Given the importance of improving small molecule induced phospholipidosis for de-risking drug discovery and improving the quality of chemical probes,[6,7] we have developed a live-cell, high-content imaging assay allowing time-dependent studies and developed robust prediction models.

Time-dependent consequences of PL on cellular signaling are poorly understood. Importantly, our experimental results demonstrated the importance of studying time dependency in live cells, to ensure the correct characterization of compounds of interest. This time dependency may also explain some inconsistencies regarding the effect of compounds to induce DIPL found in the literature, although other factors and experimental setups such as cell line of choice could also have contributed. Overall, the imaging assay that we developed has the advantage of tracking PL development across time without any dye toxicity and allowing for simultaneously assessing the influence of a compound on cell viability thus reducing costs and resources. Generating large-scale, robust datasets with experimental methods that track DIPL across time might hold the key to improve ML model performance further.

Our computational methods confirmed previous observations that ML outperforms methods relying solely on physicochemical properties.[28] RF was the best-performing algorithm in our assessment using public data as well as in our two in-house datasets. The AUC value of our model on the test set was similar to previously published methods.[28] However, when benchmarked with two external datasets, our approach outperformed FP-ADMET[39] a recently published and easy-to-implement RF to predict DIPL (Figure 4E). Given that both methods use RF, we believe that our larger compound set (aprox. 600 more compounds) for ML model construction, and the concatenation of features were key to improve performance and illustrate the advantages of our approach. Therefore, our algorithm, together with the comprehensive dataset that we have curated, represents a valuable addition to the available tool kit to predict DIPL. Using the SHAP approach[39,40] to interpret the best-performing model, it became clear that pKa and cLogP are still important features. However, many other features can also significantly contribute to the accuracy of the prediction. While pKa and cLogP values represent a property of the entire molecule, the remaining top features represent substructure properties including but not limited to cationic amphiphilic structures, such as peripheral amines and aromatic rings. However, by conducting this feature analysis on misclassified cases, we also identified compounds displaying many PL features that did not induce PL.

Divergent phospholipidosis profiles among closely related analogs that share similar physicochemical properties suggested that the mechanism for inducing phospholipidosis is complex. Comparing data on related analogs also illustrated the challenge of training computational models that reliably predict PL using reduced and simplistic physicochemical properties. On the other hand, identifying highly similar analogs that contain both inducer and non-inducer properties presented an opportunity for medicinal chemistry endeavors aimed at fine-tuning the structural features of phospholipidosis inducers to render them not inducing PL.

We identified in ~20% of cases very similar compounds displaying opposing PL phenotypes including four pairs comprising a chemical probe and its inactive analog. Chemical probes are frequently used in cell assays to investigate the association between a specific gene and a phenotype[53] and matched negative controls are expected to increase the confidence that observed phenotypes are indeed driven by the targeted protein. Our results highlight the need to also comprehensively characterize negative controls for phenotypes such as PL, as the observed phenotype might be caused by PL induction rather than the presumed target. This hypothesis is in agreement with recent reports on distinct off-target effects of probes and their negative controls.[54]

Our ML model and experimental data identified 21 chemical probes of high-quality chemical probes in the Chemical Probes Portal (https://www.chemicalprobes.org/) that induce phospholipidosis, some at the recommended cellular concentrations. Thus, the additional annotation of the compounds evaluated in the Chemical Probes Portal will add value for the biomedical community to this highly annotated compound set. With the improved PL prediction algorithm and the live-cell assay, we moreover provide a tool set to identify masked PL inducers among small molecule screening collections to increase robustness and reproducibility in biomedical research. However, closely related analogs such as chemical probes and their controls which differ in PL inductions also offer interesting tools studying the consequences of PL on cellular signaling and potentially, they will lead to PL biomarkers for drug safety studies in the future.

Limitations of the study

While the ML models developed in this study demonstrated robust PL prediction capabilities, the existence of structurally closely related compounds showing opposite PL induction behavior illustrates the limitations that still exist to comprehensively predict PL. These results not only highlight the complex nature of PL induction, which should be further explored through additional experimental studies, but also emphasizes the need to develop more advanced modeling techniques and descriptors to improve computational methods for reliable prediction of complex phenomena such as PL. For example, the pKa calculations focused only on the most basic center of each small molecule, and therefore the representation of overall molecular basicity in the context of multiprotic molecules is an area of potential improvement. Until the performance of computational methods improves further, we recommend that all computational predictions are experimentally validated, in particular for key compounds such as chemical probes and their negative controls. Since current rationale has predominantly centered around chemical patterns and physicochemical properties (a

Cell Chemical Biology
Article

                       CellPress

generalized PL mechanism), the elucidation of PL phenomena in target-based mechanisms is particularly relevant for structurally similar molecules with different PL inductions, which often share common chemical features. Disparities in off-target profiles related to lipid metabolism or synthesis among analogs could potentially serve as an explanatory avenue meriting thorough investigation. Given the differences we have identified between cell lines, care should be taken in selecting the relevant cell system to be used for such validation. Characterization of more cell lines would also be important to better understand PL thresholds in different cellular systems. In the present study, we only tested a fraction of the small molecules in the Chemical Probes Portal, and it is thus possible that some PL-inducing chemical probes were not identified.

Significance

Phospholipidosis is a pathological condition characterized by an excessive accumulation of phospholipids in lysosomes, which is frequently induced by the exposure of cells to exogenous substances, including various pharmacological agents. This DIPL can have significant implications for drug safety and side effect profiles, potentially causing toxicity and should therefore be identified early on in the screening process for small molecules. We present a comprehensive analysis of diverse compounds for their capability to induce phospholipidosis. Assessing two diverse chemical libraries in our high-content live-cell assay, we annotated DIPL over time. The results were used to identify and implement an improved state-of-the-art machine learning algorithm to annotate phospholipidosis induction of compounds effectively, extending the currently used features to predict phospholipidosis and rationalizing the prediction via using SHAP approach. By incorporating these elements, we not only validated the effectiveness of our annotation approach but also demonstrate that close analogs are able to induce divergent phospholipidosis phenotypes. We thus contribute to the knowledge and collective understanding of phospholipidosis as well as developed tools for compound annotation.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell lines and cell line culture
- METHOD DETAILS
  - Phospholipidosis detection using confocal microscopy
  - AlamarBlueTM cell viability assessment of cell staining dyes
  - Autophagy LC3 HiBiT reporter assay
  - Chemogenomic compound and lysosomal modulation dataset curation

- Literature-reported phospholipidosis data curation and structure standardization
- Molecular representations
- Building a simple physicochemical property model
- FP-ADMET model
- Chemical similarity calculation
- Machine learning models and hyperparameter optimization
- Model interpretation and probability calculation
- Matched molecular pair generation
- Chemical Probes Portal
- QUANTIFICATION AND STATISTICAL ANALYSIS

AUTHOR CONTRIBUTIONS

H.H., A.T., S.K., A.A.A., and S.M. designed the research. A.T. and S.M. performed and managed all the experimental part of the work. H.H. and A.A.A. performed and managed all the computational work. H.H., A.T., S.K., A.A.A., and S.M. conducted data analysis and interpretation. H.H., A.T., A.A.A., and S.M. wrote the manuscript with contributions from S.K.

∂ CellPress

Cell Chemical Biology
Article

REFERENCES

1. Shahane, S.A., Huang, R., Gerhold, D., Baxa, U., Austin, C.P., and Xia, M. (2014). Detection of phospholipidosis induction: a cell-based assay in high-throughput and high-content format. J. Biomol. Screen 19, 66–76. https://doi.org/10.1177/1087057113502851.

2. Breiden, B., and Sandhoff, K. (2019). Emerging mechanisms of drug-induced phospholipidosis. Biol. Chem. 401, 31–46. https://doi.org/10.1515/hsz-2019-0270.

3. Kodavanti, U.P., and Mehendale, H.M. (1990). Cationic amphiphilic drugs and phospholipid storage disorder. Pharmacol. Rev. 42, 327–354.

4. Blik, E., Orleanska, J., Mateuszuk, L., Baranska, M., Majzner, K., and Chlopicki, S. (2022). Raman and fluorescence imaging of phospholipidosis induced by cationic amphiphilic drugs in endothelial cells. Biochim. Biophys. Acta Mol. Cell Res. 1869, 119186. https://doi.org/10.1016/j.bbamcr.2021.119186.

5. Donato, M.T., and Gómez-Lechón, M.J. (2012). Drug-induced liver steatosis and phospholipidosis: cell-based assays for early screening of drug candidates. Curr. Drug Metabol. 13, 1160–1173. https://doi.org/10.2174/138920012802850001.

6. Edwards, A., and Hartung, I.V. (2021). No shortcuts to SARS-CoV-2 antivirals. Science 373, 488–489. https://doi.org/10.1126/science.abj9488.

7. Tummino, T.A., Rezelj, V.V., Fischer, B., Fischer, A., O'Meara, M.J., Monel, B., Vallet, T., White, K.M., Zhang, Z., Alon, A., et al. (2021). Drug-induced phospholipidosis confounds drug repurposing for SARS-CoV-2. Science 373, 541–547. https://doi.org/10.1126/science.abi4708.

8. Tjaden, A., Chaikuad, A., Kowarz, E., Marschalek, R., Knapp, S., Schröder, M., and Müller, S. (2022). Image-Based Annotation of Chemogenomic Libraries for Phenotypic Screening. Molecules 27, 1439. https://doi.org/10.3390/molecules27041439.

9. Hinkovska-Galcheva, V., Treadwell, T., Shillingford, J.M., Lee, A., Abe, A., Tesmer, J.J.G., and Shayman, J.A. (2021). Inhibition of lysosomal phospholipase A2 predicts drug-induced phospholipidosis. J. Lipid Res. 62, 100089. https://doi.org/10.1016/j.jlr.2021.100089.

10. Gulbins, E., Palmada, M., Reichel, M., Lüth, A., Böhmer, C., Amato, D., Müller, C.P., Tischbirek, C.H., Groemer, T.W., Tabatabai, G., et al. (2013). Acid sphingomyelinase-ceramide system mediates effects of antidepressant drugs. Nat. Med. 19, 934–938. https://doi.org/10.1038/nm.3214.

11. Gulbins, A., Schumacher, F., Becker, K.A., Wilker, B., Soddemann, M., Boldrin, F., Müller, C.P., Edwards, M.J., Goodman, M., Caldwell, C.C., et al. (2018). Antidepressants act by inducing autophagy controlled by sphingomyelin-ceramide. Mol. Psychiatr. 23, 2324–2346. https://doi.org/10.1038/s41380-018-0090-9.

12. Anderson, N., and Borlak, J. (2006). Drug-induced phospholipidosis. FEBS Lett. 580, 5533–5540. https://doi.org/10.1016/j.febslet.2006.08.061.

13. Goldman, S.D.B., Funk, R.S., Rajewski, R.A., and Krise, J.P. (2009). Mechanisms of amine accumulation in, and egress from, lysosomes. Bioanal 1, 1445–1459. https://doi.org/10.4155/bio.09.128.

14. Reasor, M.J., Hastings, K.L., and Ulrich, R.G. (2006). Drug-induced phospholipidosis: issues and future directions. Expet Opin. Drug Saf. 5, 567–583. https://doi.org/10.1517/14740338.5.4.567.

15. Joshi, U.M., Rao, P., Kodavanti, S., Lockard, V.G., and Mehendale, H.M. (1989). Fluorescence studies on binding of amphiphilic drugs to isolated lamellar bodies: relevance to phospholipidosis. Biochim. Biophys. Acta 1004, 309–320. https://doi.org/10.1016/0005-2760(89)90078-7.

16. Chatman, L.A., Morton, D., Johnson, T.O., and Anway, S.D. (2009). A strategy for risk management of drug-induced phospholipidosis. Toxicol. Pathol. 37, 997–1005. https://doi.org/10.1177/0192623309352496.

17. Reasor, M.J. (1989). A review of the biology and toxicologic implications of the induction of lysosomal lamellar bodies by drugs. Toxicol. Appl. Pharmacol. 97, 47–56. https://doi.org/10.1016/0041-008x(89)90054-9.

18. Gum, R.J., Hickman, D., Fagerland, J.A., Heindel, M.A., Gagne, G.D., Schmidt, J.M., Michaelides, M.R., Davidsen, S.K., and Ulrich, R.G. (2001). Analysis of two matrix metalloproteinase inhibitors and their me-

19. Kasahara, T., Tomita, K., Murano, H., Harada, T., Tsubakimoto, K., Ogihara, T., Ohnishi, S., and Kakinuma, C. (2006). Establishment of an in vitro high-throughput screening assay for detecting phospholipidosis-inducing potential. Toxicol. Sci. 90, 133–141. https://doi.org/10.1093/toxsci/kfj067.

20. Coleman, J., Xiang, Y., Pande, P., Shen, D., Gatica, D., and Patton, W.F. (2010). A live-cell fluorescence microplate assay suitable for monitoring vacuolation arising from drug or toxic agent treatment. J. Biomol. Screen 15, 398–405. https://doi.org/10.1177/1087057110364242.

21. Bauch, C., Bevan, S., Woodhouse, H., Dilworth, C., and Walker, P. (2015). Predicting in vivo phospholipidosis-inducing potential of drugs by a combined high content screening and in silico modelling approach. Toxicol. Vitro 29, 621–630. https://doi.org/10.1016/j.tiv.2015.01.014.

22. Ploemen, J.P.H.T.M., Kelder, J., Hafmans, T., van de Sandt, H., van Burgsteden, J.A., Salemink, P.J.M., and van Esch, E. (2004). Use of physico-chemical calculation of pKa and CLogP to predict phospholipidosis-inducing potential: a case study with structurally related piperazines. Exp. Toxicol. Pathol. 55, 347–355. https://doi.org/10.1078/0940-2993-00338.

23. Tomizawa, K., Sugano, K., Yamada, H., and Horii, I. (2006). Physicochemical and cell-based approach for early screening of phospholipidosis-inducing potential. J. Toxicol. Sci. 31, 315–324. https://doi.org/10.2131/jts.31.315.

24. Hanumegowda, U.M., Wenke, G., Regueiro-Ren, A., Yordanova, R., Corradi, J.P., and Adams, S.P. (2010). Phospholipidosis as a function of basicity, lipophilicity, and volume of distribution of compounds. Chem. Res. Toxicol. 23, 749–755. https://doi.org/10.1021/tx9003825.

25. Schieferdecker, S., Eberlein, A., Vock, E., and Bellmann, M. (2022). Development of an in silico consensus model for the prediction of the phospholipigenic potential of small molecules. Comput. Toxicol. 22, 100226. https://doi.org/10.1016/j.comtox.2022.100226.

26. Przybylak, K.R., and Cronin, M.T.D. (2011). In Silico Studies of the Relationship Between Chemical Structure and Drug Induced Phospholipidosis. Mol. Inform. 30, 415–429. https://doi.org/10.1002/minf.201000164.

27. Przybylak, K.R., Alzahrani, A.R., and Cronin, M.T.D. (2014). How does the quality of phospholipidosis data influence the predictivity of structural alerts? J. Chem. Inf. Model. 54, 2224–2232. https://doi.org/10.1021/ci500233k.

28. Sun, H., Shahane, S., Xia, M., Austin, C.P., and Huang, R. (2012). Structure based model for the prediction of phospholipidosis induction potential of small molecules. J. Chem. Inf. Model. 52, 1798–1805. https://doi.org/10.1021/ci3001875.

29. Slavov, S., Stoyanova-Slavova, I., Li, S., Zhao, J., Huang, R., Xia, M., and Beger, R. (2017). Why are most phospholipidosis inducers also hERG blockers? Arch. Toxicol. 91, 3885–3895. https://doi.org/10.1007/s00204-017-1995-9.

30. Sun, H., Xia, M., Shahane, S.A., Jadhav, A., Austin, C.P., and Huang, R. (2013). Are hERG channel blockers also phospholipidosis inducers? Bioorg. Med. Chem. Lett. 23, 4587–4590. https://doi.org/10.1016/j.bmcl.2013.06.034.

31. Fusani, L., Brown, M., Chen, H., Ahlberg, E., and Noeske, T. (2017). Predicting the Risk of Phospholipidosis with in Silico Models and an Image-Based in Vitro Screen. Mol. Pharm. 14, 4346–4352. https://doi.org/10.1021/acs.molpharmaceut.7b00388.

32. Rodríguez-Pérez, R., and Bajorath, J. (2021). Explainable Machine Learning for Property Predictions in Compound Optimization. J. Med. Chem. 64, 17744–17752. https://doi.org/10.1021/acs.jmedchem.1c01789.

33. Antolin, A.A., Sanfelice, D., Crisp, A., Villasclaras Fernandez, E., Mica, I.L., Chen, Y., Collins, I., Edwards, A., Müller, S., Al-Lazikani, B., and Workman, P. (2023). The Chemical Probes Portal: an expert review-based public resource to empower chemical probe assessment, selection and use.

Cell Chemical Biology
Article

CellPress

Nucleic Acids Res. 51, D1492–D1502. https://doi.org/10.1093/nar/gkac909.

34. Nioi, P., Perry, B.K., Wang, E.J., Gu, Y.Z., and Snyder, R.D. (2007). In vitro detection of drug-induced phospholipidosis using gene expression and fluorescent phospholipid based methodologies. Toxicol. Sci. 99, 162–173. https://doi.org/10.1093/toxsci/kfm157.

35. Steen, N.V.D., Honeywell, R.J., Dekker, H., Meerloo, J.V., Kole, J., Musters, R., Ruijtenbeek, R., Rolfo, C., Pauwels, P., Peters, G.J., and Giovannetti, E. (2018). Resistance to crizotinib in a cMET gene amplified tumor cell line is associated with impaired sequestration of crizotinib in lysosomes. J. Mol. Clin. Med. 1, 99–106. https://doi.org/10.31083/j.jmcm.2018.02.007.

36. Xue, L., and Liu, P. (2021). Daurisoline inhibits hepatocellular carcinoma progression by restraining autophagy and promoting cisplatin-induced cell death. Biochem. Biophys. Res. Commun. 534, 1083–1090. https://doi.org/10.1016/j.bbrc.2020.09.068.

37. Herp, D., Ridinger, J., Robaa, D., Shinsky, S.A., Schmidtkunz, K., Yesiloglu, T.Z., Bayer, T., Steimbach, R.R., Herbst-Gervasoni, C.J., Merz, A., et al. (2022). First Fluorescent Acetylspermidine Deacetylation Assay for HDAC10 Identifies Selective Inhibitors with Cellular Target Engagement. Chembiochem 23, e202200180. https://doi.org/10.1002/cbic.202200180.

38. Venkatraman, V. (2021). FP-ADMET: a compendium of fingerprint-based ADMET prediction models. J. Cheminf. 13, 75. https://doi.org/10.1186/s13321-021-00557-5.

39. Štrumbelj, E., and Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. Knowl. Inf. Syst. 41, 647–665. https://doi.org/10.1007/s10115-013-0679-x.

40. Shapley, L.S. (1953). 17. A Value for n-Person Games. In Contributions to the Theory of Games (AM-28), Volume II, K. Harold William and T. Albert William, eds. (Princeton University Press), pp. 307–318. https://doi.org/10.1515/9781400881970-018.

41. Muehlbacher, M., Tripal, P., Roas, F., and Kornhuber, J. (2012). Identification of drugs inducing phospholipidosis by novel in vitro data. ChemMedChem 7, 1925–1934. https://doi.org/10.1002/cmdc.201200306.

42. Liu, X., Suo, H., Zhou, S., Hou, Z., Bu, M., Liu, X., and Xu, W. (2021). Afatinib induces pro-survival autophagy and increases sensitivity to apoptosis in stem-like HNSCC cells. Cell Death Dis. 12, 728. https://doi.org/10.1038/s41419-021-04011-0.

43. Li, Z., Yang, C., Li, X., Du, X., Tao, Y., Ren, J., Fang, F., Xie, Y., Li, M., Qian, G., et al. (2020). The dual role of BI 2536, a small-molecule inhibitor that targets PLK1, in induction of apoptosis and attenuation of autophagy in neuroblastoma cells. J. Cancer 11, 3274–3287. https://doi.org/10.7150/jca.33110.

44. da Motta, L.L., Ledaki, I., Purshouse, K., Haider, S., De Bastiani, M.A., Baban, D., Morotti, M., Steers, G., Wigfield, S., Bridges, E., et al. (2017). The BET inhibitor JQ1 selectively impairs tumour response to hypoxia and downregulates CA9 and angiogenesis in triple negative breast cancer. Oncogene 36, 122–132. https://doi.org/10.1038/onc.2016.184.

45. Lansu, K., Karpiak, J., Liu, J., Huang, X.P., McCorvy, J.D., Kroeze, W.K., Che, T., Nagase, H., Carroll, F.I., Jin, J., et al. (2017). In silico design of novel probes for the atypical opioid receptor MRGPRX2. Nat. Chem. Biol. 13, 529–536. https://doi.org/10.1038/nchembio.2334.

46. Griffen, E., Leach, A.G., Robb, G.R., and Warner, D.J. (2011). Matched molecular pairs as a medicinal chemistry tool. J. Med. Chem. 54, 7739–7750. https://doi.org/10.1021/jm200452d.

47. Hussain, J., and Rea, C. (2010). Computationally efficient algorithm to identify matched molecular pairs (MMPs) in large data sets. J. Chem. Inf. Model. 50, 339–348. https://doi.org/10.1021/ci900450m.

48. Robichaud, J., Fournier, J.F., Gagné, S., Gauthier, J.Y., Hamel, M., Han, Y., Hénault, M., Kargman, S., Levesque, J.F., Mamane, Y., et al. (2011). Applying the pro-drug approach to afford highly bioavailable antagonists of P2Y(14). Bioorg. Med. Chem. Lett. 21, 4366–4368. https://doi.org/10.1016/j.bmcl.2010.12.113.

49. Eram, M.S., Shen, Y., Szewczyk, M., Wu, H., Senisterra, G., Li, F., Butler, K.V., Kaniskan, H.Ü., Speed, B.A., Dela Seña, C., et al. (2016). A Potent, Selective, and Cell-Active Inhibitor of Human Type I Protein Arginine Methyltransferases. ACS Chem. Biol. 11, 772–781. https://doi.org/10.1021/acschembio.5b00839.

50. Karim, R.M., Chan, A., Zhu, J.Y., and Schönbrunn, E. (2020). Structural Basis of Inhibitor Selectivity in the BRD7/9 Subfamily of Bromodomains. J. Med. Chem. 63, 3227–3237. https://doi.org/10.1021/acs.jmedchem.9b01980.

51. Nakayama, K., Szewczyk, M.M., Dela Sena, C., Wu, H., Dong, A., Zeng, H., Li, F., de Freitas, R.F., Eram, M.S., Schapira, M., et al. (2018). TP-064, a potent and selective small molecule inhibitor of PRMT4 for multiple myeloma. Oncotarget 9, 18480–18493. https://doi.org/10.18632/oncotarget.24883.

52. Szewczyk, M.M., Ishikawa, Y., Organ, S., Sakai, N., Li, F., Halabelian, L., Ackloo, S., Couzens, A.L., Eram, M., Dilworth, D., et al. (2020). Pharmacological inhibition of PRMT7 links arginine monomethylation to the cellular stress response. Nat. Commun. 11, 2396. https://doi.org/10.1038/s41467-020-16271-z.

53. Arrowsmith, C.H., Audia, J.E., Austin, C., Baell, J., Bennett, J., Blagg, J., Bountra, C., Brennan, P.E., Brown, P.J., Bunnage, M.E., et al. (2015). The promise and peril of chemical probes. Nat. Chem. Biol. 11, 536–541. https://doi.org/10.1038/nchembio.1867.

54. Lee, J., and Schapira, M. (2021). The Promise and Peril of Chemical Probe Negative Controls. ACS Chem. Biol. 16, 579–585. https://doi.org/10.1021/acschembio.1c00036.

55. Tjaden, A., Glessmann, R.T., Knapp, S., Schröder, M., and Müller, S. (2022). High-content live-cell multiplex screen for chemogenomic compound annotation based on nuclear morphology. STAR Protoc. 3, 101791. https://doi.org/10.1016/j.xpro.2022.101791.

56. Lazar, D.F., Gillette, A.A., Lewis, S.R., and Butler, B.L. (2021). Chapter 4 - Multimodal assessment of autophagy in mammalian cells with a novel, LC3-based tandem reporter. In Methods Cell Biol., O. Kepp and L. Galluzzi, eds. (Academic Press), pp. 39–57. https://doi.org/10.1016/bs.mcb.2021.02.004.

57. Orogo, A.M., Choi, S.S., Minnier, B.L., and Kruhlak, N.L. (2012). Construction and Consensus Performance of (Q)SAR Models for Predicting Phospholipidosis Using a Dataset of 743 Compounds. Mol. Inform. 31, 725–739. https://doi.org/10.1002/minf.201200048.

58. Goracci, L., Ceccarelli, M., Bonelli, D., and Cruciani, G. (2013). Modeling phospholipidosis induction: reliability and warnings. J. Chem. Inf. Model. 53, 1436–1446. https://doi.org/10.1021/ci400113t.

59. Pelletier, D.J., Gehlhaar, D., Tilloy-Ellul, A., Johnson, T.O., and Greene, N. (2007). Evaluation of a published in silico model and construction of a novel Bayesian model for predicting phospholipidosis inducing potential. J. Chem. Inf. Model. 47, 1196–1205. https://doi.org/10.1021/ci6004542.

60. Lowe, R., Glen, R.C., and Mitchell, J.B.O. (2010). Predicting phospholipidosis using machine learning. Mol. Pharm. 7, 1708–1714. https://doi.org/10.1021/mp100103e.

61. Dolciami, D., Villasclaras-Fernandez, E., Kannas, C., Meniconi, M., Al-Lazikani, B., and Antolin, A.A. (2022). canSAR chemistry registration and standardization pipeline. J. Cheminf. 14, 28. https://doi.org/10.1186/s13321-022-00606-7.

62. David, L., Thakkar, A., Mercado, R., and Engkvist, O. (2020). Molecular representations in AI-driven drug discovery: a review and practical guide. J. Cheminf. 12, 56. https://doi.org/10.1186/s13321-020-00460-5.

63. Durant, J.L., Leland, B.A., Henry, D.R., and Nourse, J.G. (2002). Reoptimization of MDL keys for use in drug discovery. J. Chem. Inf. Comput. Sci. 42, 1273–1280. https://doi.org/10.1021/ci010132r.

64. Rogers, D., and Hahn, M. (2010). Extended-connectivity fingerprints. J. Chem. Inf. Model. 50, 742–754. https://doi.org/10.1021/ci100050t.

65. Yap, C.W. (2011). PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J. Comput. Chem. 32, 1466–1474. https://doi.org/10.1002/jcc.21707.

66. Maggiora, G., Vogt, M., Stumpfe, D., and Bajorath, J. (2014). Molecular similarity in medicinal chemistry. J. Med. Chem. 57, 3186–3204. https://doi.org/10.1021/jm401411z.

67. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

68. Fix, E., and Hodges, J.L., Jr. (1951). Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties (Randolph Field: USAF School of Aviation Medicine). Technical Report 4.

69. Altman, N.S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. Am. Statistician *46*, 175–185. https://doi.org/10.1080/00031305.1992.10475879.

70. Breiman, L. (2001). Random Forests. Mach. Learn. *45*, 5–32. https://doi.org/10.1023/A:1010933404324.

71. Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. Ann. Stat. *7*, 1–26. https://doi.org/10.1214/aos/1176344552.

72. Cortes, C., and Vapnik, V. (1995). Support-vector networks. Mach. Learn. *20*, 273–297. https://doi.org/10.1007/BF00994018.

73. Shawe-Taylor, J., and Cristianini, N. (2004). Kernel Methods for Pattern Analysis (Cambridge University Press). https://doi.org/10.1017/CBO9780511809682.

74. Chen, T., and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Association for Computing Machinery).

75. Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. Ann. Stat. *29*, 1189–1232. https://www.jstor.org/stable/2699986.

76. Hu, X., Hu, Y., Vogt, M., Stumpfe, D., and Bajorath, J. (2012). MMP-Cliffs: systematic identification of activity cliffs on the basis of matched molecular pairs. J. Chem. Inf. Model. *52*, 1138–1145. https://doi.org/10.1021/ci3001138.

Cell Chemical Biology
Article

CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Hoechst33342	Thermo Scientific	Cat#: 62249 https://www.thermofisher.com/order/catalog/product/62249
HCS LipidTOX Red phospholipidosis detection reagent	Thermo Scientific	Cat#: H34351 https://www.thermofisher.com/order/catalog/product/H34351?SID=srch-srp-H34351
LysoTracker Green DND-26	Thermo Scientific	Cat#: L7526 https://www.thermofisher.com/order/catalog/product/de/en/L7526#:~:text=LysoTracker%20Green%20DND%2D26%20is,excited%20using%20a%20FITC%20filter.
Trypan blue 0.4%	Thermo Scientific	Cat#: 15250061 https://www.thermofisher.com/order/catalog/product/15250061
Critical commercial assays		
AlamarBlue Cell Viability Reagent	Invitrogen	Cat#: DAL1025
Autophagy LC3 HiBiT Reporter Assay	Promega	Cat#: GA1050
Experimental models: cell lines		
U2OS cells (female, 15-years old)	ATCC	HTB-96™
HEK293T (female, fetus)	ATCC	CRL-1573™
HepG2 (male, 15-years old)	ATCC	HB-8065™
Software and algorithms		
PubChemPy	GitHub	https://github.com/mcs07/PubChemPy
KNIME 4.6.2	KNIME Analytics Platform	https://www.knime.com/
canSARchem workflow	KNIME Hub	https://hub.knime.com/danieladolciami/spaces/Public/latest/canSARchem~hquSdFp3di4kiEv
RDKit version 2022.03.5	Package	https://www.rdkit.org/
PaDELPy	GitHub	https://github.com/ecrl/padelpy
MOE 2020.09	Software	https://www.chemcomp.com/Products.htm
ChemAxon/Instant JChem Version 22.16.0	Software	https://chemaxon.com/instant-jchem
scikit-learn 1.0.2	Package	https://scikit-learn.org/stable/
Python 3.9.10	Programming language	https://www.python.org/
Cytoscape version 3.9.1	Software	https://cytoscape.org/
SHAP 0.41.0	Package	https://shap.readthedocs.io/en/latest/index.html
CQ1 microscope software	Yokogawa	v 1.04.03.01
CellPathfinder Software v3.04.02.02	Yokogawa	N/A
GraphPad Prism v8.4.3	Graphpad Software	https://www.graphpad.com/features
PL prediction model	This paper	https://github.com/HuabinHu/ML-for-PL-prediction
Other		
CQ1 microscope	Yokogawa	N/A
384-well cell culture microplate, PS, f-bottom, µClear®	Greiner	Cat#: 781091
ECHO® 550 Acoustic Liquid Handler	Labcyte	N/A
ECHO® source plate	Labcyte	Cat#: P.05525

271

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cytomat2C24 incubator	Thermo Scientific	N/A
TC20 Automated Cell Counter	Bio-Rad	https://www.bio-rad.com/de-de/product/tc20-automated-cell-counter?ID=M7FBG34VY
PHERAstar® plate reader	BMG Labtech	N/A
DMEM medium plus L-glutamine (high glucose)	Thermo Scientific	Cat#11965084
McCoy's 5A medium	Gibco	Cat#15410604
Fetal bovine serum (FBS)	Thermo Scientific	Cat#26140079
Penicillin-Streptomycin (10,000 U/ml)	Thermo Scientific	Cat#14190144
DPBS	Thermo Scientific	Cat#14190-094
Geneticin™ (G418 Sulfate)	Thermo Scientific	Cat#10131035

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Susanne Müller, susanne.mueller-knapp@bmls.de.

Materials availability
This study did not generate any new reagents.

Data and code availability
- All data reported including curated dataset, chemogenomic compound set, lysosomal modulation set, and chemical probes in this paper are provided in the supplementary files and are publicly available as of the date of publication.
- The source code used in this study has been deposited at https://github.com/HuabinHu/ML-for-PL-prediction and are publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines and cell line culture
U2OS (female, 15-years old) cells were regularly tested for mycoplasma infection. Cells were grown in DMEM medium supplemented with 10% FBS and 1% Penicillin/Streptomycin (100 U/ml penicillin and 100 mg/mL streptomycin) at 37°C and 5% CO_2.

HEK293T (female, fetus) cells were regularly tested for mycoplasma infection. Cells were grown in DMEM medium supplemented with 10% FBS and 1% Penicillin/Streptomycin (100 U/ml penicillin and 100 mg/mL streptomycin) at 37°C and 5% CO_2.

HepG2 (male, 15-years old) cells were regularly tested for mycoplasma infection. Cells were grown in DMEM medium supplemented with 10% FBS and 1% Penicillin/Streptomycin (100 U/ml penicillin and 100 mg/mL streptomycin) at 37°C and 5% CO_2.

U2OS Autophagy LC3 HiBiT Reporter (female, 15-years old) cells were regularly tested for mycoplasma infection. Cells were grown in McCoy's 5A medium supplemented by 10% FBS (Thermo Scientific) and 250 μg/mL G418 (Thermo Scientific) at 37°C and 5% CO_2.

All cell lines were obtained from ATCC (details are described in the key resource table) without further authentication. Influence of sex, gender or both is not expected.

METHOD DETAILS

Phospholipidosis detection using confocal microscopy
Human Osteosarcoma cells (U2OS) (ATCC HTB-96), human hepatocellular carcinoma cells (HepG2) (ATCC HB-8065) and human embryonic kidney cells (HEK293T) (ATCC CRL-1573) were cultured in DMEM plus L-glutamine (high glucose) supplemented by 10% FBS (Thermo Scientific) and Penicillin/Streptomycin (Thermo Scientific). HCS LipidTOX Red phospholipidosis detection reagent (Thermo Scientific) was diluted in cell culture medium 1:1000 and incubated for 5 min at 37°C and 5% CO_2. To remove any aggregates, the dye was filtered using 0.2 μm sterile filters. Additionally, the cells were stained with 60 nM Hoechst33342 (Thermo Scientific) and 75 nM LysoTracker Green DND-26 (Thermo Scientific). For every cell line 1500 cells per well in 50 μL/well were seeded in 384- well plates (cell culture microplate, PS, f-bottom, μClear, 781091, Greiner) in culture medium supplemented with the indicated

Cell Chemical Biology
Article

⌀ CellPress

amount of cell staining dyes. All compounds used in this study, were tested at 10 μM over a 48 h time period in U2OS cells. Compounds from the Lysosomal modulation set and hit compounds from the chemical probe portal were additionally tested at 1 μM over a 72 h period. Reference compounds (amiodarone, chloroquin, chlorpromazin, sertralin, tamoxifen, haloperidol) were tested at three different concentrations (10 μM, 1 μM and 0.5 μM) in all three cell lines (HEK293T, HepG2 and U2OS) and were used as positive controls for all experiments. Cellular shape and fluorescence were measured at 10x magnification at 6 h, 12 h, 24 h, 48 h as well as 72 h after compound exposure using a CQ1 high-content confocal microscope (Yokogawa). Reference compounds were also imaged at 60x magnification. The following setup parameters were used for image acquisition: Ex 405 nm/Em 447/60 nm, 500 ms, 50% (Hoechst33342); Ex 561 nm/Em 617/73 nm, 100 ms, 40% (HCS LipidTOX); Ex 488/Em 525/50 nm, 50 ms, 40% (LysoTracker Green); bright field, 300 ms, 100% transmission, one centered field per well, 7 z-stacks per well with 55 μm spacing. Images were analyzed using the CellPathfinder software (Yokogawa). Viability assessment using an ML algorithm implemented in the CellPathfinder software was performed as described previously by Tjaden et al.[8,55] In brief, cell body and cell nuclei were detected and gated in different categories based on different features of the cells. Cells showing "Hoechst High Intensity Objects" were excluded. Cells defined as "Normal" were further gated in cells showing healthy, fragmented or pyknosed nuclei.[55] Phospholipidosis dots were defined as being included in lysosomes based on LysoTracker staining. Phospholipidosis was calculated using the total intensity (average) of channel 3. Data were normalized against the average of DMSO (0.1%) treated cells. All experiments were performed in biological duplicates. Significance was shown which was calculated by using an unpaired multiple t test analysis of timepoints 12 h and 48 h using GraphPad Prism 8.4.3.

AlamarBlueTM cell viability assessment of cell staining dyes

Human Osteosarcoma cells (U2OS) (ATCC HTB-96) were cultured in DMEM plus L-glutamine (high glucose) supplemented by 10% FBS (Thermo Scientific) and Penicillin/Streptomycin (Thermo Scientific). 5000 cells per well were seeded in 384-well plates (cell culture microplate, PS, f-bottom, μClear, 781091, Greiner) containing 50 μL/well culture medium one day prior to experimental readout. Cells were incubated over night at 37°C and 5% CO_2. Six different concentrations of HCS LipidTOX Red phospholipidosis detection reagent (Thermo Scientific) (1:4000, 1:2000, 1:1000, 1:500, 1:250, 1:125) or LysoTracker Green DND-26 (Thermo Scientific) (400 nM, 200 nM, 100 nM, 50 nM, 25 nM, 12.5 nM) or both (1:1000 HCD LipidTOX Red phospholipidosis detection reagent and 75 nM LysoTracker Green DND-26) were added in six replicates. Staurosporine 10 μM was added as a positive control. After 24 h of compound incubation alamarBlueTM HS was added 1:10 (10 μL/well) to every well according to the manufacturer's instructions. Plates were incubated for 1.5 h protected from light at 37°C and 5% CO_2. Fluorescence at an excitation wavelength of 560 nm and emission wavelength of 590 nm was measured using a PHERAstar plate reader (BMG Labtech). Data were normalized against the average of DMSO (0.1%) treated cells. Two biological replicates were tested.

Autophagy LC3 HiBiT reporter assay

Autophagy flux was detected as described previously.[56] In brief, U2OS Autophagy LC3 HiBiT Reporter Cells (Promega) were cultured in McCoy's 5A medium (Gibco) supplemented by 10% FBS (Thermo Scientific) and 250 μg/mL G418 (Thermo Scientific). 2000 cells per well were seeded in 384-well plates (cell culture microplate, PS, f-bottom, low volume, 784075, Greiner) containing 10 μL/well culture medium one day prior to experimental readout. Cells were incubated over night at 37°C and 5% CO_2. Compounds (Lysosomal modulation set) were added at two different concentrations (2.5 μM and 10 μM). Nano-Glo HiBiT Lytic Reagent was added to the cells 1:100 prior to readout. Luminescence was measured 6 h, 24 h and 48 h after compound exposure with or without autophagy induction of 250 nM Torin using a PHERAstar plate reader (BMG Labtech).

Chemogenomic compound and lysosomal modulation dataset curation

The "chemogenomic compound set" was selected from of our in-house compound library, including compounds covering kinases, epigenetic and GPCR targets. The screen was performed to further annotate the already implemented in-house library. The second set called "lysosomal modulation set" was selected from an in-house library of compounds, containing a collection of small molecules involved in autophagy.

Literature-reported phospholipidosis data curation and structure standardization

Following a comprehensive review of phospholipidosis-related literature, we manually compiled compounds with reported *in vivo* or *in vitro* phospholipidosis annotations from nine literature sources.[23,24,27,31,41,57–60] Compounds sourced from diverse origins were subjected to rigorous standardization protocols involving several steps. First, the compound's name was converted to SMILES-format strings with the aid of the PubChemPy package (https://github.com/mcs07/PubChemPy). Subsequently, the chemical structures of the compounds were standardized using an open-access pipeline, namely the canSARchem registration workflow.[61] This pipeline involved several steps such as salt removal, tautomerization, and charge neutralization. This multistep standardization protocol ensured the identification of duplicates, enabling appropriate subsequent analyses.

Molecular representations

In cheminformatics, various molecular representations, such as molecular fingerprints and numerical physicochemical descriptors, have been developed and widely adopted to encode chemical or physical information of small molecules in a computer-readable format for multiple applications.[62] To explore the performance of ML prediction, we employed three widely used two-dimensional

CellPress

Cell Chemical Biology
Article

(2D) fingerprints and one 2D physicochemical descriptor from Molecular Operating Environment (MOE 2020.09, https://www.chemcomp.com/Products.htm), namely: (1) the Molecular ACCess System (MACCS) fingerprint,[63] which contains 166 predefined structural patterns encoded by a fixed position with a length of 166 bits (i.e., 166 substructures); (2) the extended-connectivity fingerprint (ECFP), which is a circular topological fingerprint capturing local atom environments by considering circular layers centered at each heavy atom with increasing bond diameter up to a maximum of a predefined value.[64] Here, we employed the ECFP fingerprint with a bond diameter of 4 (ECFP4), which was folded into a 2048-bit vector; (3) the PubChem fingerprints, which encodes a total of 881 predefined molecular fragments or patterns to represent the chemical information of a molecule (https://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem_fingerprints.txt); and (4) 206 MOE 2D descriptors, which provide partial charge, pharmacophore feature information, subdivided surface areas, and other physical properties in contrast to the previously mentioned fingerprints. The calculation of MACCS fingerprint was performed using the RDKit package version 2022.03.5 (https://www.rdkit.org/). Given several existing implementations and variations for computing ECFP4 fingerprint (e.g., ChemAxon, OpenEye), we utilize the ECFP4 fingerprint implemented in RDKit (Morgan/circular fingerprint with radius of 2). PubChem fingerprint were generated the by PaDELPy library,[65] while MOE 2D descriptors were computed by MOE 2020.09 software (https://www.chemcomp.com/Products.htm). In addition, the physicochemical properties of cLogP and pKa (the most basic center) were calculated using ChemAxon/Instant JChem version 22.16.0 (https://chemaxon.com/instant-jchem). These properties were concatenated with the molecular representations mentioned above. To fit with the binary nature of the molecular fingerprints, the continuous cLogP and pKa values were transformed into binary values, with a bit set to one if the cLogP value was greater than or equal to 3.0 (or if the pKa value was greater than or equal to 7.4) and a bit set to zero if the cLogP value was less than 3.0 (or if the pKa value was less than 7.4).[41]

Building a simple physicochemical property model

Previously calculated pKa and cLogP values were used to calculate Ploemen's rule formulated as: $(pKa)^2 + (cLogP)^2 \geq 90$, in conjunction with $pKa \geq 8$ and $cLogP \geq 1$, as the criteria for classifying a compound as a potential phospholipidosis inducer.[22]

FP-ADMET model

We utilized a recently developed model (https://gitlab.com/vishsoft/fpadmet) capable of predicting absorption, distribution, metabolism, excretion, and toxicity (ADMET) for the compounds of interest.[38] This model incorporates 20 distinct binary fingerprints, enabling the modeling of over 50 ADMET-related properties. Specifically, we employed the model to assess its performance on our two external sets on PL prediction, aiming to compare it with our own developed model.

Chemical similarity calculation

Tanimoto coefficient (Tc), as a popular similarity metric, was used to quantify the structural similarity between two compounds.[66] Tc is generally defined as:

$$Tc(A, B) = \frac{c}{a + b - c}$$

where a and b are the number of features present in compounds A and B, respectively, and c is the number of features shared by A and B. Here, molecules are represented by ECFP fingerprint with radius of 2 (ECFP4) and folded into 2048-bit vector. Tc measurements were carried out in RDKit package version 2022.03.5 (https://www.rdkit.org/).

Machine learning models and hyperparameter optimization

To discriminate phospholipidosis inducers from non-inducers, four distinct ML algorithms were employed to generate classification tasks. All ML models were generated using the python-implemented scikit-learn 1.0.2 (https://scikit-learn.org/stable/).[67]

K-nearest neighbor

A k-nearest neighbor (k-NN) algorithm is a non-parametric and supervised ML method introduced by Fix and Hodges.[68] It can be used to solve both classification and regression problems. In classification tasks, the algorithm uses predefined distance metrics to calculate the ranked distances between the compound of interest (unknown sample) and its k nearest neighbors in the training dataset.[69] The majority vote (class label) among the k neighbors is assigned to the compound of interest. The value of k is a critical parameter that needs to be optimized to control the bias and variance of the method.

Random forest

A random forest (RF) classifier is a popular supervised ML algorithm that involves an ensemble of decision trees.[70] During the model training, each individual tree is constructed using a bootstrap sampling method[71] and a random subset of features for node splitting, resulting in a diverse set of trees. The prediction for a given input data is obtained by aggregating the outputs of all individual trees and determining the majority class vote. RF is known for its robustness against overfitting and its ability to handle high-dimensional datasets.

Support vector machine

Support Vector Machines (SVMs) are a class of supervised ML algorithms widely used for classification tasks.[72] The main objective of SVM is to find the optimal hyperplane(s) in a high-dimensional feature space that can optimally separate support vectors from each class, thereby maximizing the margin or distance between the classes. However, if the input data points are not linearly separable in the original feature space, a kernel function can be applied to map them into a higher-dimensional feature space where they can be

Cell Chemical Biology
Article

CellPress

linearly separable.[73] The choice of kernel function, including polynomial, radial basis function (RBF), and sigmoid kernel, is critical and often depends on the nature of the data and the classification task.

eXtreme gradient boosting

eXtreme Gradient Boosting (XGBoost) is a highly popular supervised ensemble machine learning algorithm used for classification and regression tasks.[74] Unlike RF, which relies on independent decision trees and a bagging-based algorithm, XGBoost follows a gradient boosting approach.[75] It constructs the model in a sequential stage-wise fashion by progressively adding weak learners, typically decision trees that perform slightly better than random guessing. The algorithm iteratively improves the ensemble by leveraging the prediction performance of the preceding learner. It dynamically adjusts the weight distribution, giving greater emphasis to samples that have a significant influence on the construction of the subsequent learner. By aggregating multiple shallow learners and considering the importance or weight assigned to each learner, XGBoost effectively reduces bias and enhances predictive performance. XGBoost is implemented in *xgboost* package (https://xgboost.readthedocs.io/en/stable/).

Hyperparameter optimization

The ML models were fine-tuned by adjusting their parameters through the use of an internal stratified 5-fold cross-validation technique, which was implemented using the scikit-learn package[67] version 1.0.2 (https://scikit-learn.org/stable/) implemented in python version 3.9.10.

For the *k*-NN algorithm, the optimal k values were selected from a range of [1, 3, 5, 7, 10], and the leaf size was adjusted with values of 30, 40, and 50.

For the Random Forest algorithm, the following parameters were optimized: (1) the number of decision trees ("n_estimators": 100, 200, 300, 400); (2) the number of features considered during the search for the best split ("max_features": sqrt, log2, 0.7); (3) the minimum number of samples required at a leaf node ("min_samples_leaf": 1, 3, 5, 10); (4) the maximum depth of the tree ("max_depth": 7, 10, 12, None). Additionally, the model was built using a balanced class weight.

For the SVM algorithm, the following parameters, which greatly impact the performance of the model, were fine-tuned: (1) the regularization parameter ("C": 0.1, 1, 10, 50, 100, 200, 300, 400, 500, 1000); (2) the selection of the kernel function (polynomial, RBF, and sigmoid kernel); (3) the kernel coefficient gama ("gama": 1, 0.1, 0.01, 0.001).

For XGBoost algorithm, the following parameters were optimized: (1) the step size shrinkage used for update ("learning rate": 0.05, 0.15, 0.25); (2) maximum depth of a tree ("max_depth": 3, 5, 7, 9); (3) minimum sum of instance weight (hessian) needed in a child ("min_child_weight": 1, 3, 5, 7); (4) minimum loss reduction required to make a further partition on a leaf node of the tree ("gamma": 2, 4); (5) the portion of columns to be randomly samples for individual base trees ("colsample_bytree": 0.3, 0.5); (6) the number of decision trees (n_estimators: 100, 300, 400, 500).

Performance evaluation

For all ML models, 70% of the curated phospholipidosis data were used for model training and hyperparameter optimization. The best performing hyperparameter combinations were then selected to generate the final classifier, which was trained with the complete training set. The remaining 30% of curated data (test set), coupled with two external datasets (chemogenomic compound and lysosomal modulation set), were utilized for evaluating the prediction performance of the model in a realistic scenario using 10 independent trials. To assess the model's performance, the recall, F1-score (F1), accuracy, balanced accuracy (BA), and Matthew's Correlation Coefficient (MCC) were calculated using the following equations:

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$F1 = \frac{2TP}{2TP+FP+FN}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$BA = \frac{0.5TP}{TP+FN} + \frac{0.5TN}{TN+FP}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

where TP denotes "true positives", TN "true negatives", FP "false positives", and FN "false negatives". Recall measures the model's ability to detect positive samples; F1 score indicates the harmonic mean of precision and recall; accuracy represents the overall prediction accuracy of the model; BA is the arithmetic mean of sensitivity and specificity when dealing with imbalanced data; MCC a metric used to assess the performance of binary classification models, a score of 0 indicates that the classifier performs no better

CellPress

Cell Chemical Biology
Article

than random guessing. Moreover, the Receiver Operating Characteristic (ROC) curve and its corresponding Area Under ROC (AUC) are also utilized as a threshold-independent metric to further assess the overall predictive performance. With the exception of the MCC, which ranges from −1 to 1, all other metrics are scaled between 0 and 1 with 1 representing perfect classification.

Model interpretation and probability calculation

In order to gain better insights into the underlying factors that contribute to the phospholipidosis prediction, we utilized a locally interpretable explanation method called Shapley Additive exPlanations (SHAP).[39,40] This approach, based on game theory, was originally developed to allocate the total gain among players in a cooperative team, allowing for the evaluation of individual player contributions to the outcome of a game. In the context of our phospholipidosis prediction model, Shapley values were used to estimate the contributions of each feature to the final output. The magnitude and direction (sign) of the feature contribution are manifested by the Shapley values. Positive Shapley values indicate that a feature positively contributes to the prediction of phospholipidosis inducers, whereas for features with negative contributions to inducer prediction (i.e., positive contributions to non-inducer prediction), Shapley values with a negative sign will be used. All features work in a cooperative manner, and the final output prediction reflects the additive Shapley values of all features. To calculate the predicted probability of a test compound using SHAP, the method first calculates the expected probability also referring to as the base probability. The base probability is obtained as the average model output over training set instances and corresponds to the predicted probability of a test compound with unknown feature values. Then, SHAP values are calculated for each feature of the test compound quantifying the contribution of each feature to the predicted probability relative to the expected value. The SHAP values are added to the expected value to modify the base probability. By combining the SHAP values with the base probability, we can obtain the final output probability of the ML model. SHAP analysis is carried out via using SHAP package version 0.41.0 (https://shap.readthedocs.io/en/latest/index.html).

Matched molecular pair generation

Matched molecular pair (MMP) is defined as a pair of compounds distinguished by a single site.[46] In order to identify structurally similar pairs, a computationally efficient fragmentation algorithm was introduced to systematically generate MMPs in large-scale manner. This involved systematically fragmenting exocyclic single bonds in the compounds, yielding two fragments at each step.[47] To obtain analogs that typically generated during compound optimization, strict size constraints were implemented during the fragmentation process to derive a core and substituent fragment. In particular, the core fragment was mandated to consist of no fewer than 1.5 times the number of non-hydrogen atoms present in the substituent fragment. Furthermore, the maximal size of fragment substituent was set to 13 non-hydrogen atoms.[76] All MMPs were generated using MMP-related nodes that were integrated into KNIME 4.6.2 (https://www.knime.com/). Following the generation of MMPs, they were organized in a network format using Cytoscape version 3.9.1 (https://cytoscape.org/).

Chemical Probes Portal

Compounds from the Chemical Probes Portal (https://www.chemicalprobes.org/), a publicly available resource based on expert reviews that aims to enable the assessment, selection, and utilization of chemical probes,[33] were downloaded (accessed in August, 2022). Only the compounds with an in-cell rating of three stars or more were selected, resulting in 321 high-quality chemical probes (from 795 probes in total). The best machine learning model was used to predict if any of them could potentially induce phospholipidosis. After removing those already existing in the curated dataset, we further prioritized the probes with a predicted probability value greater than or equal to 0.7 for experimental validation, yielding 31 probes for experimental confirmation.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analysis was performed using GraphPad Prism 8.4.3. Statistical details and definition of parameters are stated in the STAR methods sections and figure legends. Sample sizes are indicated in the respective figure legend.

Appendix J: List of publications

Publications highlighted in blue are part of the dissertation (9).

<u>First-author publications</u>

1. **Tjaden, A.**; Chaikuad, A.; Kowarz, E.; Marschalek, R.; Knapp, S.; Schröder, M.; Müller, S., Image-Based Annotation of Chemogenomic Libraries for Phenotypic Screening. Molecules 2022, 27 (4), 1439.

> Contribution Menge, A.: Conceptualization, methodology and formal analysis, validation, data curation, writing—original draft preparation, writing, review and editing.

2. **Tjaden, A**.; Giessmann, R. T.; Knapp, S.; Schröder, M.; Müller, S., High-content live-cell multiplex screen for chemogenomic compound annotation based on nuclear morphology. STAR Protocols 2022, 3 (4), 101791.

> Contribution Menge, A.: Conceptualization, Methodology, Investigation, Validation and Formal Analysis, writing – Original Draft, writing, review and editing.

3. Hu, H. *; **Tjaden, A. ***; Knapp, S.; Antolin, A. A.; Müller, S., A machine learning and live-cell imaging tool kit uncovers small molecules induced phospholipidosis. *Cell Chemical Biology* 2023.

> Contribution Menge, A.: Conceptualization, Methodology, Investigation, Validation and Formal Analysis, writing – Original Draft, writing, review and editing.

*shared first author

4. Rak, M. *; **Tjaden, A. *:** Chemogenomic tool compounds targeting mammalian STE20-like protein kinases 1-4 (MST1-4) and development of a MST3 selective inhibitor

> Contribution Menge, A.: Conceptualization, Investigation, Validation and Formal Analysis, writing – Original Draft, Writing – Review & Editing

→ to be submitted soon

*shared first author

Book contributions

1. **Tjaden, A.**; Knapp, S.; Müller, S., Annotation of the Effect of Chemogenomic Compounds on Cell Health Using High-Content Microscopy in Live-Cell Mode. In *Chemogenomics: Methods and Protocols*, Merk, D.; Chaikuad, A., Eds. Springer US: New York, NY, 2023; pp 59-73.

> Contribution Menge, A.: Conceptualization, Methodology, Investigation, Validation and Formal Analysis, writing – Original Draft, writing, review and editing.

2. Elson, L.; **Tjaden, A.**; Knapp, S.; Müller, S., Characterization of Cellular Viability Using Label-Free Brightfield Live-Cell Imaging. In *Chemogenomics: Methods and Protocols*, Merk, D.; Chaikuad, A., Eds. Springer US: New York, NY, 2023; pp 75-88.

> Contribution Menge, A.: Conceptualization, Original Draft, Writing – Review & Editing

Co-author publications

1. Rak, M.; Tesch, R.; Berger, L. M.; Shevchenko, E.; Raab, M.; **Tjaden, A.**; Zhubi, R.; Balourdas, D.-I.; Joerger, A. C.; Poso, A.; Krämer, A.; Elson, L.; Lučić, A.; Kronenberger, T.; Hanke, T.; Strebhardt, K.; Sanhaji, M.; Knapp, S., Shifting the selectivity of pyrido[2,3-d]pyrimidin-7(8H)-one inhibitors towards the salt-inducible kinase (SIK) subfamily. *European Journal of Medicinal Chemistry* **2023,** *254*, 115347.

> Contribution Menge, A.: Viability assessment using Multiplex high-content live-cell assay, contribution to study design, data evaluation, figure preparation, manuscript editing.

2. Tredup, C.; Ndreshkjana, B.; Schneider, N. S.; **Tjaden, A.;** Kemas, A. M.; Youhanna, S.; Lauschke, V. M.; Berger, B.-T.; Krämer, A.; Berger, L. M.; Röhm, S.; Knapp, S.; Farin, H. F.; Müller, S., Deep Annotation of Donated Chemical Probes (DCP) in Organotypic Human Liver Cultures and Patient-Derived Organoids from Tumor and Normal Colorectum. *ACS Chemical Biology* **2023**.Contribution Tjaden, A.: Viability assessment using Multiplex high-content live-cell assay, contribution to study design, data evaluation, figure preparation, manuscript editing.

> Contribution Menge, A.: Viability assessment using Multiplex high-content live-cell assay, contribution to study design, data evaluation, figure preparation, manuscript editing.

3. Amrhein, J. A.; Berger, L. M.; **Tjaden, A.**; Krämer, A.; Elson, L.; Tolvanen, T.; Martinez-Molina, D.; Kaiser, A.; Schubert-Zsilavecz, M.; Müller, S.; Knapp, S.; Hanke, T., Discovery of 3-Amino-1H-pyrazole-Based Kinase Inhibitors to Illuminate the Understudied PCTAIRE Family. International Journal of Molecular Sciences 2022, 23 (23), 14834.

> Contribution Menge, A.: Conducting fluorescent ubiquitination-based cell cycle reporter (FUCCI) assay, contribution to study design, data evaluation, figure preparation, manuscript writing and editing.

4. Krämer, A.; **Tjaden, A.**; Ndreshkjana, B.; Tredup, C.; Farin, H. F.; Knapp, S.; Janin, Y. L.; Müller, S., IPP/CNRS-A017: A chemical probe for human dihydroorotate dehydrogenase (hDHODH). Current Research in Chemical Biology 2022, 2, 100034.

> Contribution Menge, A.: Viability assessment using Multiplex high-content live-cell assay, Uridine rescue experiment in U2OS cells using brightfield microscopy, contribution to study design, data evaluation, figure preparation, manuscript editing.

5. Ehrler, J. H. M.; Brunst, S.; **Tjaden, A.**; Kilu, W.; Heering, J.; Hernandez-Olmos, V.; Krommes, A.; Kramer, J. S.; Steinhilber, D.; Schubert-Zsilavecz, M.; Müller, S.; Merk, D.; Proschak, E., Compilation and evaluation of a fatty acid mimetics screening library. Biochemical Pharmacology 2022, 204, 115191.

> Contribution Menge, A.: Viability assessment using Multiplex high-content live-cell assay, data evaluation, figure preparation, manuscript editing.

6. Kurz, C. G.; Preuss, F.; **Tjaden, A.**; Cusack, M.; Amrhein, J. A.; Chatterjee, D.; Mathea, S.; Berger, L. M.; Berger, B.-T.; Krämer, A.; Weller, M.; Weiss, T.; Müller, S.; Knapp, S.; Hanke, T., Illuminating the Dark: Highly Selective Inhibition of Serine/Threonine Kinase 17A with Pyrazolo[1,5-a]pyrimidine-Based Macrocycles. Journal of Medicinal Chemistry 2022, 65 (11), 7799-7817.

> Contribution Menge, A.: Viability assessment using Multiplex high-content live-cell assay, data evaluation, figure preparation, manuscript editing.

7. Röhm, S.; Berger, B.-T.; Schröder, M.; Chatterjee, D.; Mathea, S.; Joerger, A. C.; Pinkas, D. M.; Bufton, J. C.; **Tjaden, A.**; Kovooru, L.; Kudolo, M.; Pohl, C.; Bullock, A. N.; Müller, S.; Laufer, S.; Knapp, S., Development of a Selective Dual Discoidin Domain Receptor (DDR)/p38 Kinase Chemical Probe. Journal of Medicinal Chemistry 2021, 64 (18), 13451-13474.

> Contribution Menge, A.: Viability assessment using Multiplex high-content live-cell assay, data evaluation, figure preparation, manuscript editing.

8. Wells, C. I.; Drewry, D. H.; Pickett, J. E.; **Tjaden, A.**; Krämer, A.; Müller, S.; Gyenis, L.; Menyhart, D.; Litchfield, D. W.; Knapp, S.; Axtman, A. D., Development of a potent and selective chemical probe for the pleiotropic kinase CK2. Cell Chemical Biology 2021, 28 (4), 546-558.e10.

> Contribution Menge, A.: Western blot analysis and quantification of SGC-CK-2-1 compound, data evaluation, figure preparation, manuscript editing.

9. Heitel, P.; Faudone, G.; Helmstädter, M.; Schmidt, J.; Kaiser, **A.; Tjaden**, A.; Schröder, M.; Müller, S.; Schierle, S.; Pollinger, J.; Merk, D., A triple farnesoid X receptor and peroxisome proliferator-activated receptor α/δ activator reverses hepatic fibrosis in diet-induced NASH in mice. Communications Chemistry 2020, 3 (1), 174.

> Contribution Menge, A.: Viability assessment using Multiplex high-content live-cell assay, data evaluation, figure preparation, manuscript editing.

10. Krämer, A.; Kurz, C. G.; Berger, B.-T.; Celik, I. E.; **Tjaden, A.**; Greco, F. A.; Knapp, S.; Hanke, T., Optimization of pyrazolo[1,5-a]pyrimidines lead to the identification of a highly selective casein kinase 2 inhibitor. European Journal of Medicinal Chemistry 2020, 208, 112770.

> Contribution Menge, A.: Western blot analysis and quantification of SGC-CK-2-1 compound, data evaluation, figure preparation, manuscript editing.

Appendix K: List of Abbreviations

Abbreviation   Full name

Abbreviation	Full name
2D	Two-dimensional
3D	Three-dimensional
AI	Artificial intelligence
ALK	Anaplastic lymphoma kinase
ALK5	Activin receptor-like kinase 5
APD	Action potential duration
ATP	Adenosine triphosphate
AURKA	Aurora kinase A
BRAF	v-raf murine sarcoma viral oncogene homolog B1
BRD	Bromodomain
BrdU	5-bromo-2'-deoxyuridine
CADs	Cationic amphiphilic drugs
CDKs	Cyclin-dependent kinases
Cdt1	Chromatin Licensing and DNA Replication Factor 1
CK2	Casein kinase 2
cMET	Mesenchymal-epithelial transition factor
COVID-19	Coronavirus disease of 2019
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DAPK	Death-associated protein kinase
DCP	Donated Chemical Probes
DDR	dual Discoidin domain receptor
DED	Death-inducing domain
Diablo	Direct IAP binding protein with low PI
DIPL	Drug-induced phopsholipidosis
DISC	Death-inducing signal complex
DNA	Deoxyribonucleic acid
DR	Death receptor
DRAK1	Death-associated protein kinase-related apoptosis-inducing protein kinase 1
EdU	5-ethynyl-2'-deoxyuridine
EMBL	European Molecular Biology Laboratory
EMBL-EBI	European Molecular Biology Laboratory (EMBL) European Bioinformatics Institute's
ER	Endoplasmic reticulum
EUbOPEN	Enabling and Unlocking Biology in the OPEN
FDA	Food and Drug Administration
FUCCI	Fluorescent Ubiquitination-based Cell Cycle Indicator
G1-phase	Gap 1 phase

GFP	Green fluorescent protein
GPCR	G-protein-coupled receptors
GR	Growth rate
GSK	Glycogen synthase kinase
HCA	High-content analysis
HCI	High-content imaging
HCS	High-content screening
HCT	Human colorectal cancer cells
HEK293T	Human embryonic kidney cells
HepG2	Hepatoblastoma cells
HtrA2/Omi	High-temperature requirement family member HtrA2/Omi
IAPs	Inhibitors of apoptosis proteins
IMI	Innovative medicine initiative
KNN	K-nearest neighbor
LAMP	Lysosomal-associated membrane protein
LIMK1	LIM domain kinase 1
MCM	Mini chromosome maintenance
ML	Machine learning
MLKL	Mixed lineage kinase domain like pseudokinase
MOMP	Mitochondrial outer membrane permeabilization
MPT	Mitochondrial permeability transition
MRC	Human fibroblasts
MTH1	Mut-T homolog-1
mTOR	Mammalian target of rapamycin
MTT	3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NBD-PE	1-acyl-2-(12[(7-nitro-2-1-,3-benzoxadiazol-4-yl)amino]dodecanoyl
NEK11	NIMA-related kinase 11
NFκB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NSCLC	Non-small cell lung cancer
PAINS	Pan-assay interference
PCNA	Proliferating cell nuclear antigen
PKC	Protein kinase C
PL	Phospholipidosis
PLA2G15	Lysosome-specific phospholipase A2
RF	Random forest
RFP	Red fluorescent protein
RIPK	Receptor-interacting proteins
RNA	Ribonucleic acid
ROS	Reactive oxygen species
ROS1	ROS proto-oncogene 1
RTK	Receptor tyrosine kinase
S phase	DNA synthesis phase
SEM	Standard error of mean
SGC	Structural Genomics Consortium
SIK1	Salt-inducible kinase 1
Smac	Second mitochondria derived activator of caspases

STK17A	Serine/threonine kinase 17A
SVM	Support vector machine
TNF	Tumor necrosis factor
TNFR1	Tumor necrosis factor receptor 1
TRADD	TNFR1-associated death domain protein
U2OS	Osteosarcoma cells
ULK1	Unc-51 like autophagy activating kinase 1
VEGF	Vascular endothelial growth factor
XGBoost	EXtreme Gradient Boosting
XTT	2,3-bis[2-methoxy-4-nitro-5-sulfophenyl]-2H-tetrazolium-5-carboxanilide

7. References

1. Ziegler, S., Pries, V., Hedberg, C. & Waldmann, H. Target Identification for Small Bioactive Molecules: Finding the Needle in the Haystack. *Angewandte Chemie International Edition* **52**, 2744-2792 (2013).

2. Oh, E.S. & Petronis, A. Origins of human disease: the chrono-epigenetic perspective. *Nature Reviews Genetics* **22**, 533-546 (2021).

3. Anderson, S. *Making medicines: a brief history of pharmacy and pharmaceuticals*. (Pharmaceutical Press, 2005).

4. Mohs, R.C. & Greig, N.H. Drug discovery and development: Role of basic biological research. *Alzheimer's & Dementia: Translational Research & Clinical Interventions* **3**, 651-657 (2017).

5. Sterling, J., Baker, J.R., McCluskey, A. & Munoz, L. Systematic literature review reveals suboptimal use of chemical probes in cell-based biomedical research. *Nature Communications* **14**, 3228 (2023).

6. Bredel, M. & Jacoby, E. Chemogenomics: an emerging strategy for rapid target and drug discovery. *Nature Reviews Genetics* **5**, 262-275 (2004).

7. Drewes, G. & Knapp, S. Chemoproteomics and Chemical Probes for Target Discovery. *Trends in Biotechnology* **36**, 1275-1286 (2018).

8. Frye, S.V. The art of the chemical probe. *Nature Chemical Biology* **6**, 159-161 (2010).

9. Antolin, A.A. *et al.* The Chemical Probes Portal: an expert review-based public resource to empower chemical probe assessment, selection and use. *Nucleic Acids Research* **51**, D1492-D1502 (2023).

10. Arrowsmith, C.H. *et al.* The promise and peril of chemical probes. *Nat Chem Biol* **11**, 536-541 (2015).

11. Baell, J.B. & Holloway, G.A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *Journal of Medicinal Chemistry* **53**, 2719-2740 (2010).

12. Baell, J.B. & Nissink, J.W.M. Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017-Utility and Limitations. *ACS Chem Biol* **13**, 36-44 (2018).

13. Schenone, M., Dančík, V., Wagner, B.K. & Clemons, P.A. Target identification and mechanism of action in chemical biology and drug discovery. *Nature Chemical Biology* **9**, 232-240 (2013).

14. Brunden, K.R., Trojanowski, J.Q., Smith, A.B., Lee, V.M.Y. & Ballatore, C. Microtubule-stabilizing agents as potential therapeutics for

neurodegenerative disease. *Bioorganic & Medicinal Chemistry* **22**, 5040-5049 (2014).

15. Márquez-Jurado, S. *et al.* Mitochondrial levels determine variability in cell death by modulating apoptotic gene expression. *Nat Commun* **9**, 389 (2018).

16. Tredup, C. *et al.* Deep Annotation of Donated Chemical Probes (DCP) in Organotypic Human Liver Cultures and Patient-Derived Organoids from Tumor and Normal Colorectum. *ACS Chemical Biology* (2023).

17. Sakaue-Sawano, A. *et al.* Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. *Cell* **132**, 487-498 (2008).

18. Yano, S. & Hoffman, R.M. Real-Time Determination of the Cell-Cycle Position of Individual Cells within Live Tumors Using FUCCI Cell-Cycle Imaging. *Cells* **7**, 168 (2018).

19. Zielke, N. & Edgar, B.A. FUCCI sensors: powerful new tools for analysis of cell proliferation. *WIREs Developmental Biology* **4**, 469-487 (2015).

20. Kodavanti, U.P. & Mehendale, H.M. Cationic amphiphilic drugs and phospholipid storage disorder. *Pharmacological Reviews* **42**, 327-354 (1990).

21. Shahane, S.A. *et al.* Detection of phospholipidosis induction: a cell-based assay in high-throughput and high-content format. *J Biomol Screen* **19**, 66-76 (2014).

22. DiMasi, J.A., Feldman, L., Seckler, A. & Wilson, A. Trends in Risks Associated With New Drug Development: Success Rates for Investigational Drugs. *Clinical Pharmacology & Therapeutics* **87**, 272-277 (2010).

23. Zeng, X., Tu, X., Liu, Y., Fu, X. & Su, Y. Toward better drug discovery with knowledge graph. *Current Opinion in Structural Biology* **72**, 114-126 (2022).

24. Yamaguchi, S., Kaneko, M. & Narukawa, M. Approval success rates of drug candidates based on target, action, modality, application, and their combinations. *Clin Transl Sci* **14**, 1113-1122 (2021).

25. Moffat, J.G., Vincent, F., Lee, J.A., Eder, J. & Prunotto, M. Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nature Reviews Drug Discovery* **16**, 531-543 (2017).

26. Hay, M., Thomas, D.W., Craighead, J.L., Economides, C. & Rosenthal, J. Clinical development success rates for investigational drugs. *Nature Biotechnology* **32**, 40-51 (2014).

27. Moffat, J.G., Rudolph, J. & Bailey, D. Phenotypic screening in cancer drug discovery — past, present and future. *Nature Reviews Drug Discovery* **13**, 588-602 (2014).

28. Swinney, D.C. & Anthony, J. How were new medicines discovered? *Nature Reviews Drug Discovery* **10**, 507-519 (2011).

29. Frank Simione, J. Key Issues Relating to the Genetic Stability and Preservation of Cells and Cell Banks. *PDA Journal of Pharmaceutical Science and Technology* **46**, 226-232 (1992).

30. Ranga, A. *et al.* 3D niche microarrays for systems-level analyses of cell fate. *Nature Communications* **5**, 4324 (2014).

31. Yamada, K.M. & Cukierman, E. Modeling Tissue Morphogenesis and Cancer in 3D. *Cell* **130**, 601-610 (2007).

32. Zhao, S. & Iyengar, R. Systems pharmacology: network analysis to identify multiscale mechanisms of drug action. *Annu Rev Pharmacol Toxicol* **52**, 505-521 (2012).

33. Bunnage, M.E., Chekler, E.L.P. & Jones, L.H. Target validation using chemical probes. *Nature Chemical Biology* **9**, 195-199 (2013).

34. Knapp, S. *et al.* A public-private partnership to unlock the untargeted kinome. *Nature Chemical Biology* **9**, 3-6 (2013).

35. Caron, P.R. *et al.* Chemogenomic approaches to drug discovery. *Curr Opin Chem Biol* **5**, 464-470 (2001).

36. Menzies, A.M., Long, G.V. & Murali, R. Dabrafenib and its potential for the treatment of metastatic melanoma. *Drug Design, Development and Therapy* **6**, 391-405 (2012).

37. Puszkiel, A. *et al.* Clinical Pharmacokinetics and Pharmacodynamics of Dabrafenib. *Clinical Pharmacokinetics* **58**, 451-467 (2019).

38. Klaeger, S. *et al.* The target landscape of clinical kinase drugs. *Science* **358** (2017).

39. Moret, N. *et al.* Cheminformatics Tools for Analyzing and Designing Optimized Small-Molecule Collections and Libraries. *Cell Chem Biol* **26**, 765-777.e763 (2019).

40. Poon, C.C. & Kelly, J.J. Development of crizotinib, a rationally designed tyrosine kinase inhibitor for non-small cell lung cancer. *Int J Cancer* **140**, 1945-1954 (2017).

41. Kazandjian, D. *et al.* FDA approval summary: crizotinib for the treatment of metastatic non-small cell lung cancer with anaplastic lymphoma kinase rearrangements. *Oncologist* **19**, e5-11 (2014).

42. Lineweaver, H. & Burk, D. The determination of enzyme dissociation constants. *Journal of the American chemical society* **56**, 658-666 (1934).

43. Müller, S. *et al.* Donated chemical probes for open science. *eLife* **7**, e34311 (2018).

44. Blagg, J. & Workman, P. Choose and Use Your Chemical Probe Wisely to Explore Cancer Biology. *Cancer Cell* **32**, 9-25 (2017).

45. Ahmad, G. & Amiji, M. Use of CRISPR/Cas9 gene-editing tools for developing models in drug discovery. *Drug Discovery Today* **23**, 519-533 (2018).

46. Kurata, M., Yamamoto, K., Moriarity, B.S., Kitagawa, M. & Largaespada, D.A. CRISPR/Cas9 library screening for drug target discovery. *Journal of Human Genetics* **63**, 179-186 (2018).

47. Edwards, A.M., Bountra, C., Kerr, D.J. & Willson, T.M. Open access chemical and clinical probes to support drug discovery. *Nature chemical biology* **5**, 436-440 (2009).

48. Wagner, B.K. Introduction to Chemical Probes, in *The Discovery and Utility of Chemical Probes in Target Discovery*. (eds. P. Brennan & S.V. Rodriguez) 0 (The Royal Society of Chemistry, 2020).

49. Müller, S. *et al.* Target 2035 – update on the quest for a probe for every protein. *RSC Medicinal Chemistry* **13**, 13-21 (2022).

50. Drewry, D.H. *et al.* Progress towards a public chemogenomic set for protein kinases and a call for contributions. *PLoS One* **12**, e0181585 (2017).

51. Zhang, J., Yang, P.L. & Gray, N.S. Targeting cancer with small molecule kinase inhibitors. *Nature Reviews Cancer* **9**, 28-39 (2009).

52. Chakravorty, S.J. *et al.* Nuisance Compounds, PAINS Filters, and Dark Chemical Matter in the GSK HTS Collection. *SLAS Discov* **23**, 532-545 (2018).

53. Schenone, M., Dančík, V., Wagner, B.K. & Clemons, P.A. Target identification and mechanism of action in chemical biology and drug discovery. *Nat Chem Biol* **9**, 232-240 (2013).

54. Fang, L., Pan, J. & Yan, B. High-throughput determination of identity, purity, and quantity of combinatorial library members using LC/MS/UV/ELSD. *Biotechnology and Bioengineering* **71**, 162-171 (2000).

55. Yan, B. *et al.* Quality Control in Combinatorial Chemistry: Determination of the Quantity, Purity, and Quantitative Purity of Compounds in Combinatorial Libraries. *Journal of Combinatorial Chemistry* **5**, 547-559 (2003).

56. Coimbra, M. *et al.* Improving solubility and chemical stability of natural compounds for medicinal use by incorporation into liposomes. *International Journal of Pharmaceutics* **416**, 433-442 (2011).

57. Li, X. *et al.* DeepChemStable: Chemical Stability Prediction with an Attention-Based Graph Convolution Network. *Journal of Chemical Information and Modeling* **59**, 1044-1049 (2019).

58. Lipinski, C.A., Lombardo, F., Dominy, B.W. & Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* **46**, 3-26 (2001).

59. Eastman, A. Improving anticancer drug development begins with cell culture: misinformation perpetrated by the misuse of cytotoxicity assays. *Oncotarget* **8**, 8854-8866 (2017).

60. Gayvert, K.M., Madhukar, N.S. & Elemento, O. A Data-Driven Approach to Predicting Successes and Failures of Clinical Trials. *Cell Chem Biol* **23**, 1294-1301 (2016).

61. Liu, Y., Peterson, D.A., Kimura, H. & Schubert, D. Mechanism of Cellular 3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide (MTT) Reduction. *Journal of Neurochemistry* **69**, 581-593 (1997).

62. Hansen, J. & Bross, P. A Cellular Viability Assay to Monitor Drug Toxicity, in *Protein Misfolding and Cellular Stress in Disease and Aging: Concepts and Protocols*. (eds. P. Bross & N. Gregersen) 303-311 (Humana Press, Totowa, NJ; 2010).

63. Wang, S., Yu, H. & Wickliffe, J.K. Limitation of the MTT and XTT assays for measuring cell viability due to superoxide formation induced by nano-scale TiO2. *Toxicology in Vitro* **25**, 2147-2151 (2011).

64. Berridge, M.V., Herst, P.M. & Tan, A.S. Tetrazolium dyes as tools in cell biology: New insights into their cellular reduction, in *Biotechnology Annual Review*, Vol. 11 127-152 (Elsevier, 2005).

65. Roehm, N.W., Rodgers, G.H., Hatfield, S.M. & Glasebrook, A.L. An improved colorimetric assay for cell proliferation and viability utilizing the tetrazolium salt XTT. *Journal of Immunological Methods* **142**, 257-265 (1991).

66. Lim, S.W., Loh, H.S., Ting, K.N., Bradshaw, T.D. & Allaudin, Z.N. Reduction of MTT to Purple Formazan by Vitamin E Isomers in the Absence of Cells. *Trop Life Sci Res* **26**, 111-120 (2015).

67. Bonnier, F. *et al.* Cell viability assessment using the Alamar blue assay: A comparison of 2D and 3D cell culture models. *Toxicology in Vitro* **29**, 124-131 (2015).

68. Kuda, T. & Yano, T. Colorimetric alamarBlue assay as a bacterial concentration and spoilage index of marine foods. *Food Control* **14**, 455-461 (2003).

69. Sherry, L.V.-H., Andrew, O.B., Beverly, W., Carlton, H.L. & Stephen, F.B. Application and Evaluation of the Alamarblue Assay for Cell Growth and Survival of Fibroblasts. *In Vitro Cellular & Developmental Biology. Animal* **34**, 239-246 (1998).

70. O'Brien, J., Wilson, I., Orton, T. & Pognan, F. Investigation of the Alamar Blue (resazurin) fluorescent dye for the assessment of mammalian cell cytotoxicity. *European Journal of Biochemistry* **267**, 5421-5426 (2000).

71. Malinowski, P. *et al.* Comparison of the Usefulness of MTT and CellTiterGlo Tests Applied for Cytotoxicity Evaluation of Compounds from the Group of Polyphenols. *Environmental Sciences Proceedings* **18**, 9 (2022).

72. Hannah, R., Beck, M., Moravec, R. & Riss, T. CellTiter-Glo™ Luminescent cell viability assay: a sensitive and rapid method for determining cell viability. *Promega Cell Notes* **2**, 11-13 (2001).

73.	Peternel, L., Kotnik, M., Prezelj, A. & Urleb, U. Comparison of 3 Cytotoxicity Screening Assays and Their Application to the Selection of Novel Antibacterial Hits. *SLAS Discovery* **14**, 142-150 (2009).

74.	Howarth, A. *et al.* HighVia-A Flexible Live-Cell High-Content Screening Pipeline to Assess Cellular Toxicity. *SLAS Discov* **25**, 801-811 (2020).

75.	O'Brien, P.J. High-Content Analysis in Toxicology: Screening Substances for Human Toxicity Potential, Elucidating Subcellular Mechanisms and In Vivo Use as Translational Safety Biomarkers. *Basic & Clinical Pharmacology & Toxicology* **115**, 4-17 (2014).

76.	Hanahan, D. & Weinberg, Robert A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646-674 (2011).

77.	D'Arcy, M.S. Cell death: a review of the major forms of apoptosis, necrosis and autophagy. *Cell Biology International* **43**, 582-592 (2019).

78.	Niles, A.L., Moravec, R.A. & Riss, T.L. Update on in vitro cytotoxicity assays for drug development. *Expert Opinion on Drug Discovery* **3**, 655-669 (2008).

79.	Poon, I.K.H., Lucas, C.D., Rossi, A.G. & Ravichandran, K.S. Apoptotic cell clearance: basic biology and therapeutic potential. *Nature Reviews Immunology* **14**, 166-180 (2014).

80.	Reed, J.C. Mechanisms of Apoptosis. *The American Journal of Pathology* **157**, 1415-1430 (2000).

81.	Martinvalet, D., Zhu, P. & Lieberman, J. Granzyme A Induces Caspase-Independent Mitochondrial Damage, a Required First Step for Apoptosis. *Immunity* **22**, 355-370 (2005).

82.	Li, J. & Yuan, J. Caspases in apoptosis and beyond. *Oncogene* **27**, 6194-6206 (2008).

83.	Kaufmann, S.H. *et al.* Apoptosis-associated caspase activation assays. *Methods* **44**, 262-272 (2008).

84.	Bertheloot, D., Latz, E. & Franklin, B.S. Necroptosis, pyroptosis and apoptosis: an intricate game of cell death. *Cellular & Molecular Immunology* **18**, 1106-1121 (2021).

85.	Bock, F.J. & Tait, S.W.G. Mitochondria as multifaceted regulators of cell death. *Nature Reviews Molecular Cell Biology* **21**, 85-100 (2020).

86.	Nirmala, J.G. & Lopus, M. Cell death mechanisms in eukaryotes. *Cell Biology and Toxicology* **36**, 145-164 (2020).

87.	Yan, J., Wan, P., Choksi, S. & Liu, Z.-G. Necroptosis and tumor progression. *Trends in Cancer* **8**, 21-27 (2022).

88.	Glick, D., Barth, S. & Macleod, K.F. Autophagy: cellular and molecular mechanisms. *The Journal of Pathology* **221**, 3-12 (2010).

89.	Galluzzi, L. *et al.* Molecular definitions of autophagy and related processes. *The EMBO Journal* **36**, 1811-1836 (2017).

90. Yim, W.W.-Y. & Mizushima, N. Lysosome biology in autophagy. *Cell Discovery* **6**, 6 (2020).

91. Saito, T. & Sadoshima, J. Molecular mechanisms of mitochondrial autophagy/mitophagy in the heart. *Circ Res* **116**, 1477-1490 (2015).

92. Wang, L., Klionsky, D.J. & Shen, H.-M. The emerging mechanisms and functions of microautophagy. *Nature Reviews Molecular Cell Biology* **24**, 186-203 (2023).

93. Dikic, I. & Elazar, Z. Mechanism and medical implications of mammalian autophagy. *Nature Reviews Molecular Cell Biology* **19**, 349-364 (2018).

94. Liang, X.H. *et al.* Induction of autophagy and inhibition of tumorigenesis by beclin 1. *Nature* **402**, 672-676 (1999).

95. Pasquier, B. Autophagy inhibitors. *Cellular and Molecular Life Sciences* **73**, 985-1001 (2016).

96. Touret, F. & de Lamballerie, X. Of chloroquine and COVID-19. *Antiviral Research* **177**, 104762 (2020).

97. Cortegiani, A., Ingoglia, G., Ippolito, M., Giarratano, A. & Einav, S. A systematic review on the efficacy and safety of chloroquine for the treatment of COVID-19. *Journal of Critical Care* **57**, 279-283 (2020).

98. Park, D. *et al.* Resveratrol induces autophagy by directly inhibiting mTOR through ATP competition. *Scientific Reports* **6**, 21772 (2016).

99. Dumont, F.J. & Su, Q. Mechanism of action of the immunosuppressant rapamycin. *Life Sciences* **58**, 373-395 (1995).

100. Li, J., Kim, Sang G. & Blenis, J. Rapamycin: One Drug, Many Effects. *Cell Metabolism* **19**, 373-379 (2014).

101. Asadzadeh, Z. *et al.* Current Approaches for Combination Therapy of Cancer: The Role of Immunogenic Cell Death. *Cancers* **12**, 1047 (2020).

102. Matthews, H.K., Bertoli, C. & de Bruin, R.A.M. Cell cycle control in cancer. *Nature Reviews Molecular Cell Biology* **23**, 74-88 (2022).

103. Israels, E.D. & Israels, L.G. The Cell Cycle. *The Oncologist* **5**, 510-513 (2000).

104. Fisher, R.P. The CDK Network: Linking Cycles of Cell Division and Gene Expression. *Genes Cancer* **3**, 731-738 (2012).

105. Malumbres, M. Cyclin-dependent kinases. *Genome Biology* **15**, 122 (2014).

106. Simmons Kovacs, L.A., Orlando, D.A. & Haase, S.B. Transcription networks and cyclin/CDKs: the yin and yang of cell cycle oscillators. *Cell Cycle* **7**, 2626-2629 (2008).

107. Otto, T. & Sicinski, P. Cell cycle proteins as promising targets in cancer therapy. *Nature Reviews Cancer* **17**, 93-115 (2017).

108. Eastman, A. Improving anticancer drug development begins with cell culture: misinformation perpetrated by the misuse of cytotoxicity assays. *Oncotarget* **8** (2016).

109. Schönenberger, F., Deutzmann, A., Ferrando-May, E. & Merhof, D. Discrimination of cell cycle phases in PCNA-immunolabeled cells. *BMC Bioinformatics* **16**, 180 (2015).

110. Duque, A. & Rakic, P. Different effects of bromodeoxyuridine and [3H]thymidine incorporation into DNA on cell proliferation, position, and fate. *J Neurosci* **31**, 15205-15217 (2011).

111. Seigneurin, D. & Guillaud, P. [Ki-67 antigen, a cell cycle and tumor growth marker]. *Pathol Biol (Paris)* **39**, 1020-1028 (1991).

112. Nishitani, H., Lygerou, Z. & Nishimoto, T. Proteolysis of DNA Replication Licensing Factor Cdt1 in S-phase Is Performed Independently of Geminin through Its N-terminal Region*. *Journal of Biological Chemistry* **279**, 30807-30816 (2004).

113. Singh, A.M. *et al.* Cell-cycle control of developmentally regulated transcription factors accounts for heterogeneity in human pluripotent cells. *Stem cell reports* **1**, 532-544 (2013).

114. Tjaden, A. *et al.* Image-Based Annotation of Chemogenomic Libraries for Phenotypic Screening. *Molecules* **27**, 1439 (2022).

115. Saitou, T. & Imamura, T. Quantitative imaging with Fucci and mathematics to uncover temporal dynamics of cell cycle progression. *Development, Growth & Differentiation* **58**, 6-15 (2016).

116. Koh, S.-B. *et al.* A quantitative FastFUCCI assay defines cell cycle dynamics at a single-cell level. *Journal of Cell Science* **130**, 512-520 (2017).

117. Goldstein, J.C., Waterhouse, N.J., Juin, P., Evan, G.I. & Green, D.R. The coordinate release of cytochrome c during apoptosis is rapid, complete and kinetically invariant. *Nat Cell Biol* **2**, 156-162 (2000).

118. Rehm, M., Düssmann, H. & Prehn, J.H. Real-time single cell analysis of Smac/DIABLO release during apoptosis. *J Cell Biol* **162**, 1031-1043 (2003).

119. Gad, H. *et al.* MTH1 inhibition eradicates cancer by preventing sanitation of the dNTP pool. *Nature* **508**, 215-221 (2014).

120. Gul, N. *et al.* The MTH1 inhibitor TH588 is a microtubule-modulating agent that eliminates cancer cells by activating the mitotic surveillance pathway. *Scientific Reports* **9**, 14667 (2019).

121. Breiden, B. & Sandhoff, K. Emerging mechanisms of drug-induced phospholipidosis. *Biological Chemistry* **401**, 31-46 (2020).

122. Hinkovska-Galcheva, V. *et al.* Inhibition of lysosomal phospholipase A2 predicts drug-induced phospholipidosis. *Journal of Lipid Research* **62** (2021).

123. Gulbins, E. *et al.* Acid sphingomyelinase–ceramide system mediates effects of antidepressant drugs. *Nature Medicine* **19**, 934-938 (2013).

124. Gulbins, A. *et al.* Antidepressants act by inducing autophagy controlled by sphingomyelin–ceramide. *Molecular Psychiatry* **23**, 2324-2346 (2018).

125. Anderson, N. & Borlak, J. Drug-induced phospholipidosis. *FEBS Lett* **580**, 5533-5540 (2006).

126. Goldman, S.D., Funk, R.S., Rajewski, R.A. & Krise, J.P. Mechanisms of amine accumulation in, and egress from, lysosomes. *Bioanalysis* **1**, 1445-1459 (2009).

127. Tainaka, K. *et al.* Design Strategies of Fluorescent Biosensors Based on Biological Macromolecular Receptors. *Sensors* **10**, 1355-1376 (2010).

128. Zhu, H., Fan, J., Du, J. & Peng, X. Fluorescent Probes for Sensing and Imaging within Specific Cellular Organelles. *Accounts of Chemical Research* **49**, 2115-2126 (2016).

129. Buchser, W. *et al.* Assay development guidelines for image-based high content screening, high content analysis and high content imaging. *Assay guidance manual [Internet]* (2014).

130. Ueno, T. & Nagano, T. Fluorescent probes for sensing and imaging. *Nat Methods* **8**, 642-645 (2011).

131. Liu, X. & Chang, Y.-T. Fluorescent probe strategy for live cell distinction. *Chemical Society Reviews* **51**, 1573-1591 (2022).

132. Zhang, R. *et al.* Reaction-free and MMP-independent fluorescent probes for long-term mitochondria visualization and tracking. *Chemical Science* **10**, 1994-2000 (2019).

133. Abraham, V.C., Taylor, D.L. & Haskins, J.R. High content screening applied to large-scale cell biology. *Trends in biotechnology* **22**, 15-22 (2004).

134. Stephens, D.J. & Allan, V.J. Light Microscopy Techniques for Live Cell Imaging. *Science* **300**, 82-86 (2003).

135. Tjaden, A., Giessmann, R.T., Knapp, S., Schröder, M. & Müller, S. High-content live-cell multiplex screen for chemogenomic compound annotation based on nuclear morphology. *STAR Protocols* **3**, 101791 (2022).

136. Elson, L., Tjaden, A., Knapp, S. & Müller, S. Characterization of Cellular Viability Using Label-Free Brightfield Live-Cell Imaging, in *Chemogenomics: Methods and Protocols*. (eds. D. Merk & A. Chaikuad) 75-88 (Springer US, New York, NY; 2023).

137. Tjaden, A., Knapp, S. & Müller, S. Annotation of the Effect of Chemogenomic Compounds on Cell Health Using High-Content Microscopy in Live-Cell Mode, in *Chemogenomics: Methods and Protocols*. (eds. D. Merk & A. Chaikuad) 59-73 (Springer US, New York, NY; 2023).

138. Hafner, M., Niepel, M., Chung, M. & Sorger, P.K. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nat Methods* **13**, 521-527 (2016).

139. Davis, M.I. *et al.* Comprehensive analysis of kinase inhibitor selectivity. *Nature Biotechnology* **29**, 1046-1051 (2011).

140. Kurz, C.G. *et al.* Illuminating the Dark: Highly Selective Inhibition of Serine/Threonine Kinase 17A with Pyrazolo[1,5-a]pyrimidine-Based Macrocycles. *Journal of Medicinal Chemistry* **65**, 7799-7817 (2022).

141. Schafferhans, A., O'Donoghue, S.I., Heinzinger, M. & Rost, B. Dark Proteins Important for Cellular Function. *PROTEOMICS* **18**, 1800227 (2018).

142. Essegian, D., Khurana, R., Stathias, V. & Schürer, S.C. The Clinical Kinase Index: A Method to Prioritize Understudied Kinases as Drug Targets for the Treatment of Cancer. *Cell Reports Medicine* **1**, 100128 (2020).

143. Berginski, M.E. *et al.* The Dark Kinase Knowledgebase: an online compendium of knowledge and experimental results of understudied kinases. *Nucleic Acids Research* **49**, D529-D535 (2020).

144. Vella, V., Giamas, G. & Ditsiou, A. Diving into the dark kinome: lessons learned from LMTK3. *Cancer Gene Therapy* **29**, 1077-1079 (2022).

145. Inbal, B., Bialik, S., Sabanay, I., Shani, G. & Kimchi, A. DAP kinase and DRP-1 mediate membrane blebbing and the formation of autophagic vesicles during programmed cell death. *Journal of Cell Biology* **157**, 455-468 (2002).

146. Chen, C.-H. *et al.* Bidirectional signals transduced by DAPK–ERK interaction promote the apoptotic effect of DAPK. *The EMBO Journal* **24**, 294-304 (2005).

147. Krämer, A. *et al.* Optimization of pyrazolo[1,5-a]pyrimidines lead to the identification of a highly selective casein kinase 2 inhibitor. *European Journal of Medicinal Chemistry* **208**, 112770 (2020).

148. Amrhein, J.A., Knapp, S. & Hanke, T. Synthetic Opportunities and Challenges for Macrocyclic Kinase Inhibitors. *Journal of Medicinal Chemistry* **64**, 7991-8009 (2021).

149. Mao, P. *et al.* Serine/Threonine Kinase 17A Is a Novel Candidate for Therapeutic Targeting in Glioblastoma. *PLOS ONE* **8**, e81803 (2013).

150. Röhm, S. *et al.* Development of a Selective Dual Discoidin Domain Receptor (DDR)/p38 Kinase Chemical Probe. *Journal of Medicinal Chemistry* **64**, 13451-13474 (2021).

151. Li, Y., Lu, X., Ren, X. & Ding, K. Small Molecule Discoidin Domain Receptor Kinase Inhibitors and Potential Medical Applications. *Journal of Medicinal Chemistry* **58**, 3287-3301 (2015).

152. Petersen, L.K. *et al.* Novel p38α MAP kinase inhibitors identified from yoctoReactor DNA-encoded small molecule library. *MedChemComm* **7**, 1332-1339 (2016).

153. Malumbres, M. & Barbacid, M. Cell cycle, CDKs and cancer: a changing paradigm. *Nature Reviews Cancer* **9**, 153-166 (2009).

154. Mikolcevic, P. *et al.* Cyclin-Dependent Kinase 16/PCTAIRE Kinase 1 Is Activated by Cyclin Y and Is Essential for Spermatogenesis. *Molecular and Cellular Biology* **32**, 868-879 (2012).

155. Yanagi, T. & Matsuzawa, S.-i. PCTAIRE1/PCTK1/CDK16: a new oncotarget? *Cell Cycle* **14**, 463-464 (2015).

156. Amrhein, J.A. *et al.* Discovery of 3-Amino-1H-pyrazole-Based Kinase Inhibitors to Illuminate the Understudied PCTAIRE Family. *International Journal of Molecular Sciences* **23**, 14834 (2022).

157. Charrasse, S., Carena, I., Hagmann, J., Woods-Cook, K. & Ferrari, S. PCTAIRE-1: Characterization, Subcellular Distribution, and Cell Cycle-dependent Kinase Activity. *Cell Growth & Differentiation* **10**, 611-620 (1999).

158. Yanagi, T. *et al.* PCTAIRE1/CDK16/PCTK1 is overexpressed in cutaneous squamous cell carcinoma and regulates p27 stability and cell cycle. *J Dermatol Sci* **86**, 149-157 (2017).

159. Brasca, M.G. *et al.* Identification of N,1,4,4-Tetramethyl-8-{[4-(4-methylpiperazin-1-yl)phenyl]amino}-4,5-dihydro-1H-pyrazolo[4,3-h]quinazoline-3-carboxamide (PHA-848125), a Potent, Orally Available Cyclin Dependent Kinase Inhibitor. *Journal of Medicinal Chemistry* **52**, 5152-5163 (2009).

160. Tummino, T.A. *et al.* Drug-induced phospholipidosis confounds drug repurposing for SARS-CoV-2. *Science* **373**, 541-547 (2021).

161. Bik, E. *et al.* Raman and fluorescence imaging of phospholipidosis induced by cationic amphiphilic drugs in endothelial cells. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1869**, 119186 (2022).

162. Hu, H., Tjaden, A., Knapp, S., Antolin, A.A. & Müller, S. A machine learning and live-cell imaging tool kit uncovers small molecules induced phospholipidosis. *Cell Chemical Biology* (2023).

163. Ploemen, J.-P.H.T.M. *et al.* Use of physicochemical calculation of pKa and CLogP to predict phospholipidosis-inducing potential: A case study with structurally related piperazines. *Experimental and Toxicologic Pathology* **55**, 347-355 (2004).

164. Tomizawa, K., Sugano, K., Yamada, H. & Horii, I. Physicochemical and cell-based approach for early screening of phospholipidosis-inducing potential. *J Toxicol Sci* **31**, 315-324 (2006).

165. Hanumegowda, U.M. *et al.* Phospholipidosis as a function of basicity, lipophilicity, and volume of distribution of compounds. *Chem Res Toxicol* **23**, 749-755 (2010).

166.   Reasor, M.J., Hastings, K.L. & Ulrich, R.G. Drug-induced phospholipidosis: issues and future directions. *Expert Opin Drug Saf* **5**, 567-583 (2006).

167.   Gum, R.J. *et al.* Analysis of two matrix metalloproteinase inhibitors and their metabolites for induction of phospholipidosis in rat and human hepatocytes(1). *Biochem Pharmacol* **62**, 1661-1673 (2001).

168.   Kasahara, T. *et al.* Establishment of an in vitro high-throughput screening assay for detecting phospholipidosis-inducing potential. *Toxicol Sci* **90**, 133-141 (2006).

169.   Coleman, J. *et al.* A live-cell fluorescence microplate assay suitable for monitoring vacuolation arising from drug or toxic agent treatment. *Journal of biomolecular screening* **15**, 398-405 (2010).

170.   Nioi, P., Perry, B.K., Wang, E.J., Gu, Y.Z. & Snyder, R.D. In vitro detection of drug-induced phospholipidosis using gene expression and fluorescent phospholipid based methodologies. *Toxicol Sci* **99**, 162-173 (2007).

171.   Chikte, S., Panchal, N. & Warnes, G. Use of LysoTracker dyes: A flow cytometric study of autophagy. *Cytometry Part A* **85**, 169-178 (2014).

172.   Licciardello, M.P. & Workman, P. The era of high-quality chemical probes. *RSC Med Chem* **13**, 1446-1459 (2022).

173.   Adan, A., Kiraz, Y. & Baran, Y. Cell Proliferation and Cytotoxicity Assays. *Curr Pharm Biotechnol* **17**, 1213-1221 (2016).

174.   Single, A., Beetham, H., Telford, B.J., Guilford, P. & Chen, A. A Comparison of Real-Time and Endpoint Cell Viability Assays for Improved Synthetic Lethal Drug Validation. *SLAS Discovery* **20**, 1286-1293 (2015).

175.   Boutros, M., Heigwer, F. & Laufer, C. Microscopy-Based High-Content Screening. *Cell* **163**, 1314-1325 (2015).

176.   Du, R., Huang, C., Liu, K., Li, X. & Dong, Z. Targeting AURKA in Cancer: molecular mechanisms and opportunities for Cancer therapy. *Molecular Cancer* **20**, 15 (2021).

177.   Valiathan, R.R., Marco, M., Leitinger, B., Kleer, C.G. & Fridman, R. Discoidin domain receptor tyrosine kinases: new players in cancer progression. *Cancer and Metastasis Reviews* **31**, 295-321 (2012).

178.   Graff, J.R. *et al.* The Protein Kinase Cβ–Selective Inhibitor, Enzastaurin (LY317615.HCl), Suppresses Signaling through the AKT Pathway, Induces Apoptosis, and Suppresses Growth of Human Colon Cancer and Glioblastoma Xenografts. *Cancer Research* **65**, 7462-7469 (2005).

179.   Ogbomo, H. *et al.* The anti-tumoral drug enzastaurin inhibits natural killer cell cytotoxicity via activation of glycogen synthase kinase-3β. *Biochemical Pharmacology* **81**, 251-258 (2011).

180.   Somadder, P.D. *et al.* Drug Repurposing and Systems Biology approaches of Enzastaurin can target potential biomarkers and critical pathways in Colorectal Cancer. *Computers in Biology and Medicine* **155**, 106630 (2023).

181. Graff, J.R. *et al.* The protein kinase Cbeta-selective inhibitor, Enzastaurin (LY317615.HCl), suppresses signaling through the AKT pathway, induces apoptosis, and suppresses growth of human colon cancer and glioblastoma xenografts. *Cancer Res* **65**, 7462-7469 (2005).

182. Zhang, H.-q. *et al.* Enzastaurin cardiotoxicity: QT interval prolongation, negative inotropic responses and negative chronotropic action. *Biochemical Pharmacology* **209**, 115443 (2023).

183. Lee, J. & Schapira, M. The Promise and Peril of Chemical Probe Negative Controls. *ACS Chemical Biology* **16**, 579-585 (2021).

184. Uhlenbrock, N. *et al.* Structural and chemical insights into the covalent-allosteric inhibition of the protein kinase Akt. *Chem Sci* **10**, 3573-3585 (2019).

185. Weisner, J. *et al.* Preclinical Efficacy of Covalent-Allosteric AKT Inhibitor Borussertib in Combination with Trametinib in KRAS-Mutant Pancreatic and Colorectal Cancer. *Cancer Res* **79**, 2367-2378 (2019).

186. Plangger, A. *et al.* Cytotoxicity of combinations of the pan-KRAS SOS1 inhibitor BAY-293 against pancreatic cancer cell lines. *Discover Oncology* **13**, 84 (2022).

187. Chen, A.Y., Yu, C., Bodley, A., Peng, L.F. & Liu, L.F. A New Mammalian DNA Topoisomerase I Poison Hoechst 33342: Cytotoxicity and Drug Resistance in Human Cell Cultures1. *Cancer Research* **53**, 1332-1337 (1993).

188. Durand, R.E. & Olive, P.L. Cytotoxicity, Mutagenicity and DNA damage by Hoechst 33342. *Journal of Histochemistry & Cytochemistry* **30**, 111-116 (1982).

189. Kane, R.C., Bross, P.F., Farrell, A.T. & Pazdur, R. Velcade: U.S. FDA approval for the treatment of multiple myeloma progressing on prior therapy. *Oncologist* **8**, 508-513 (2003).

190. Kane, R.C. *et al.* Bortezomib for the treatment of mantle cell lymphoma. *Clin Cancer Res* **13**, 5291-5294 (2007).

191. Mujtaba, T. & Dou, Q.P. Advances in the understanding of mechanisms and therapeutic use of bortezomib. *Discov Med* **12**, 471-480 (2011).

192. Jackson, G., Einsele, H., Moreau, P. & Miguel, J.S. Bortezomib, a novel proteasome inhibitor, in the treatment of hematologic malignancies. *Cancer Treat Rev* **31**, 591-602 (2005).

193. Anchoori, Ravi K. *et al.* A bis-Benzylidine Piperidone Targeting Proteasome Ubiquitin Receptor RPN13/ADRM1 as a Therapy for Cancer. *Cancer Cell* **24**, 791-805 (2013).

194. Dickson, P. *et al.* Physical and functional analysis of the putative Rpn13 inhibitor RA190. *Cell chemical biology* **27**, 1371-1382. e1376 (2020).

195. Stanetty, P., Röhrling, J., Schnürch, M. & Mihovilovic, M.D. Synthesis of analogs of the phenylamino-pyrimidine type protein kinase C inhibitor CGP 60474 utilizing a Negishi cross-coupling strategy. *Tetrahedron* **62**, 2380-2387 (2006).

196. Jorda, R. *et al.* How Selective Are Pharmacological Inhibitors of Cell-Cycle-Regulating Cyclin-Dependent Kinases? *Journal of Medicinal Chemistry* **61**, 9105-9120 (2018).

197. Panknin, O. *et al.* Discovery and Characterization of BAY 1214784, an Orally Available Spiroindoline Derivative Acting as a Potent and Selective Antagonist of the Human Gonadotropin-Releasing Hormone Receptor as Proven in a First-In-Human Study in Postmenopausal Women. *Journal of Medicinal Chemistry* **63**, 11854-11881 (2020).

198. Bjelobaba, I., Stojilkovic, S.S. & Naor, Z. Editorial: Gonadotropin-Releasing Hormone Receptor Signaling and Functions. *Front Endocrinol (Lausanne)* **9**, 143 (2018).

199. Dobkin-Bekman, M. *et al.* A preformed signaling complex mediates GnRH-activated ERK phosphorylation of paxillin and FAK at focal adhesions in L beta T2 gonadotrope cells. *Mol Endocrinol* **23**, 1850-1864 (2009).

200. Pal, D. *et al.* Indazole-based microtubule-targeting agents as potential candidates for anticancer drugs discovery. *Bioorganic Chemistry* **122**, 105735 (2022).

201. Eli, S., Castagna, R., Mapelli, M. & Parisini, E. Recent Approaches to the Identification of Novel Microtubule-Targeting Agents. *Frontiers in Molecular Biosciences* **9** (2022).

202. LaFrance, B.J. *et al.* Structural transitions in the GTP cap visualized by cryo-electron microscopy of catalytically inactive microtubules. *Proceedings of the National Academy of Sciences* **119**, e2114994119 (2022).

203. Gustafsdottir, S.M. *et al.* Multiplex Cytological Profiling Assay to Measure Diverse Cellular States. *PLOS ONE* **8**, e80999 (2013).

204. Bray, M.-A. *et al.* Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nature Protocols* **11**, 1757-1774 (2016).

205. Cimini, B.A. *et al.* Optimizing the Cell Painting assay for image-based profiling. *Nature Protocols* **18**, 1981-2013 (2023).

206. Akbarzadeh, M. *et al.* Morphological profiling by means of the Cell Painting assay enables identification of tubulin-targeting compounds. *Cell Chem Biol* **29**, 1053-1064.e1053 (2022).

207. Zang, R., Li, D., Tang, I.-C., Wang, J. & Yang, S.-T. Cell-based assays in high-throughput screening for drug discovery. *International Journal of Biotechnology for Wellness Industries* **1**, 31 (2012).

208. Ross, D.T. *et al.* Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genetics* **24**, 227-235 (2000).

209. Duval, K. *et al.* Modeling Physiological Events in 2D vs. 3D Cell Culture. *Physiology* **32**, 266-277 (2017).

210. Kapałczyńska, M. *et al.* 2D and 3D cell cultures—a comparison of different types of cancer cell cultures. *Archives of Medical Science* **14**, 910-919 (2018).

211. Jensen, C. & Teng, Y. Is It Time to Start Transitioning From 2D to 3D Cell Culture? *Frontiers in Molecular Biosciences* **7** (2020).

212. Yokogawa CellPathfinder software.

Declaration for work:

Except where stated otherwise by reference or acknowledgment, the work presented was generated by myself under the supervision of my advisors during my doctoral studies. Whenever a figure, table or text is identical to a previous publication, it is stated explicitly in the thesis that copyright permission and/or co-author agreement has been obtained.