Rational Inhibitor Development for Epigenetic Reader Domains

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt beim Fachbereich 14

der Johann Wolfgang Goethe - Universität

in Frankfurt am Main

von

David Jonas Heidenreich

aus Gießen

Frankfurt am Main, 2019

(D 30)

vom Fachbereich 14 der

Johann Wolfgang Goethe - Universität als Dissertation angenommen.

Dekan: Prof. Dr. C. Glaubitz

Gutachter: Prof. Dr. Stefan Knapp, Prof. Dr. Eugen Proschak

Datum der Disputation: 19.12.2019

Table of Contents

Declaration

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. All contributions from colleagues are explicitly referenced in the thesis. The material listed below was obtained in the context of collaborative research:

- **Figure 16**: VM: Virtual screening, hydration analysis and Figures 16A+16C, DSF and ITC experiments except 16C+16D; EM lab: Compound synthesis; CT: X-ray crystallography; DH+MW: DSF and ITC experiments in 16C+16D, DH: generation of Figures 16C+16D
- **Figure 17**: EW: Synthesis of compound B3; DH: generation of Figure
- **Figure 18**: EW: Synthesis of compounds and ALK^{F1174L} and PLK1 IC_{50} values in 18A+18B; DH: BRD4 DSF and ITC data in 18B, BRD4 and ALK NanoBRET data in 18C+18D, DSF Bromodomain selectivity data in 18E, generation of Figure
- **Figure 19**: EW: Synthesis of compounds B1 and B3; DH: Co-crystallisation of BRD4(1)-B1 and BRD4(1)-B3, generation of Figure
- **Figure 23**: JS: Synthesis of compound C12; DH: Compound selection, Co-crystallisation of ENL-C12 and ENL-C19, generation of Figure
- **Figure 24**: MM: Synthesis of compounds C20-C24; DH: ITC data, Co-crystallisation of ENL-C20, generation of Figure
- **Figure 25**: MM: Synthesis of compounds; DH: generation of Figure
- **Figure 26**: MM: Synthesis of compounds SGC-iMLLT and C94; DH: Co-crystallisation of ENL-SGC-iMLLT and ENL-C94, generation of Figure
- **Supplementary Table 1**: LD: Expression and Purification of BRD3 and BRDT BRDs; FP: Expression and Purification of EP300, PCAF, SP100, TRIM33B BRDs; DH: Subcloning of ENL construct, Expression and Purification of the remaining proteins
- **Supplementary Figure S4**: JS: Synthesis of compound C12; DH: Compound selection, generation of Figure
- **Supplementary Figure S6**: MM: Synthesis of compounds C21-C24; DH: ITC data, generation of Figure

DH: D. Heidenreich (Goethe University Frankfurt), MW: M. Wanior (Goethe University Frankfurt), VM: V. Myrianthopoulos (University of Athens), EM: E. Mikros (University of Athens), CT: C. Tallant (University of Oxford), EW: E. Watts (ICR London), MM: M. Moustakim (University of Oxford), JS: J. Schmidt (Goethe University Frankfurt), LD: L. A. Dutra (São Paulo State University), FP: F. Preuß (Goethe University Frankfurt)

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. The following parts of the thesis have been previously published:

- **Figures 1, 6, 8, 11**: Copyright permission obtained from Elsevier
- **Figure 2**: Figure made by co-author Suzanne Ackloo, agreement obtained
- **Figure 10**: Copyright permission obtained from Taylor & Francis
- **Figure 28**: Experiments and Figure made by co-authors Kiran Nakka and Jeffrey Dilworth, agreement obtained

Publications

1) **A structure-based approach towards identification of inhibitory fragments for eleven-nineteen-leukemia protein (ENL)**
    - Link: https://www.ncbi.nlm.nih.gov/pubmed/30407816
    - Authors: **Heidenreich D.**, Moustakim M., Schmidt J., Merk D., Brennan P.E., Fedorov O., Chaikuad A., Knapp S.
    - Status: accepted on 08/12/18 at *Journal of medicinal chemistry*
    - Contribution: **First author**, all experiments performed and all Figures generated

2) **Discovery of an MLLT1/3 YEATS Domain Chemical Probe.**
    - Link: https://www.ncbi.nlm.nih.gov/pubmed/30288907
    - Authors: Moustakim M., Christott T., Monteiro O.P., Bennett J., Giroud C., Ward J., Rogers C.M., Smith P., Panagakou I., Díaz-Sáez L., Felce S.L., Gamble V., Gileadi C., Halidi N., **Heidenreich D.**, Chaikuad A., Knapp S., Huber K.V.M., Farnie G., Heer J., Manevski N., Poda G., Al-Awar R., Dixon D.J., Brennan P.E., Fedorov O.
    - Status: accepted on 21/10/18 at *Angewandte Chemie*
    - Contribution: Co-crystal structures of ENL with SGC-iMLLT and 94 in Figures 4 and S10

3) **Designing dual inhibitors of ALK and BRD4 by tuning kinase selectivity**
    - Link: https://www.ncbi.nlm.nih.gov/pubmed/30789735
    - Authors: Watts E., **Heidenreich D.**, Tucker E., Raab M., Strebhardt K., Chesler L., Knapp S., Bellenie B., Hoelder S.
    - Status: accepted on 21/02/19 at *Journal of medicinal chemistry*
    - Contribution: *In vitro* assay data for BRD4 (DSF, ITC), NanoBRET data for BRD4 and ALK as well as co-crystal structures of BRD4 with 16i and 16k (data in Tables 1 and 2, Figures 4, 5a and 7, SI Tables 1-4 and 6)

4) **A chemical toolbox for the study of bromodomains and epigenetic signaling**
   - Link: https://www.ncbi.nlm.nih.gov/pubmed/31015424
   - Authors: Wu Q. **Heidenreich D.**, Zhou S., Ackloo S., Krämer A., Nakka K., Lima-Fernandes E., Deblois G., Duan S., Dilworth J., Lupien M., Brennan P. E., Arrowsmith C. H., Mueller S., Fedorov O., Filippakopoulos P., Knapp S.
   - Status: accepted on 23/04/19 at *Nature Communications*
   - Contribution: Data research on BRD inhibitors, Figures 1 and 3

5) **Dietary Compound Resveratrol Is a Pan-BET Bromodomain Inhibitor.**
   - Link: https://www.ncbi.nlm.nih.gov/pubmed/29077030
   - Authors: Dutra L.A., **Heidenreich D.**, Silva G.D.B.D., Man Chin C., Knapp S., Santos J.L.D.
   - Status: accepted on 20/10/17 at *Nutrients*
   - Contribution: Supervision and discussion of measurements in Figure 1

6) **Discovery and Optimization of a Selective Ligand for the Switch/Sucrose Nonfermenting-Related Bromodomains of Polybromo Protein-1 by the Use of Virtual Screening and Hydration Analysis.**
   - Link: https://www.ncbi.nlm.nih.gov/pubmed/27617704
   - Authors: Myrianthopoulos V., Gaboriaud-Kolar N., Tallant C., Hall M.L., Grigoriou S., Brownlee P.M., Fedorov O., Rogers C., **Heidenreich D.**, Wanior M., Drosos N., Mexia N., Savitsky P., Bagratuni T., Kastritis E., Terpos E., Filippakopoulos P., Müller S., Skaltsounis A.L., Downs J.A., Knapp S., Mikros E.
   - Status: accepted on 12/09/16 at *Journal of medicinal chemistry*
   - Contribution: DSF and ITC data for inhibitors 9-12 in Figure 6

7) **Selective Targeting of Bromodomains of the Bromodomain-PHD Fingers Family Impairs Osteoclast Differentiation.**
   - Link: https://www.ncbi.nlm.nih.gov/pubmed/28849908
   - Authors: Meier J.C., Tallant C., Fedorov O., Witwicka H., Hwang S.Y., van Stiphout R.G., Lambert J.P., Rogers C., Yapp C., Gerstenberger B.S., Fedele V., Savitsky P., **Heidenreich D.**, Daniels D.L., Owen D.R., Fish P.V., Igoe N.M., Bayle E.D., Haendler B., Oppermann U.C.T., Buffa F., Brennan P.E., Müller S., Gingras A.C., Odgren P.R., Birnbaum M.J., Knapp S.
   - Status: accepted on 29/08/17 at *ACS Chemical Biology*
   - Contribution: ITC data for off-targets BRD4(1) and BRD4(2) on page 5 of SI

Zusammenfassung (lang)

In eukaryotischen Zellen liegt DNA in Form von kompaktem Chromatin vor, welches aus Wiederholungen von Nukleosomen und Linker-DNA besteht. Nukleosomen werden aus jeweils zwei Kopien von Histon 2A, 2B, 3 und 4 – dem Histon-Oktamer - und dem darum gewickelten DNA-Strang gebildet. Die Wissenschaft der Epigenetik befasst sich mit allen Faktoren, die unmittelbar die Aktivität von Genen beeinflusst. Zu den wichtigsten Faktoren gehören DNA-Methylierung und post-translationale Modifikationen von N-terminalen Histonenden (*1*). Conrad Hal Waddington prägte für die Gesamtheit der epigenetischen Modifikationen im Jahr 1957 den Begriff der "epigenetischen Landschaft". Zellen durchlaufen in diesem Modell verschiedene Entwicklungsstufen, die von epigenetischen Modifikationen beeinflusst werden, bis sie einen bestimmten Phänotyp erreichen. Auch äußere Faktoren und Metaboliten beeinflussen die epigenetische Landschaft einer Zelle. Die Phänotypen beinhalten auch kranke Phänotypen der Zelle, weswegen sie für medizinische Anwendungen von hoher Bedeutung sind (*2, 3*).

Diese Dissertation setzt einen Fokus auf Histon-Acetylierung, die sowohl mit Aktivierung als auch mit Deaktivierung von Genen in Verbindung gebracht wird. Das Hinzufügen und Entfernen solcher Modifikationen ist ein stets dynamischer Prozess, der von „Schreibern" und „Löschern" (engl. *writers* und *erasers*) vollzogen wird. Acetylierung von Lysin-ε-Aminogruppen an N-terminalen Histonenden wird von sog. Histon-Acetyltransferasen (HAT) katalysiert, während der gegenteilige Schritt von sog. Histon-Deacetylasen katalysiert wird (*4, 5*). „Leser" (engl. *readers*) sind Gegenstand dieser Thesis. Sie zeigen Spezifität gegenüber einer oder mehreren Histonmodifikationen und vermitteln als Volllängenproteine den Kontakt zwischen Modifikationen, Gerüstproteinen, Transkriptionsfaktoren und RNA-Polymerase II. Acetyl-Lysine an Histonenden werden im Menschen von Bromodomänen (abgeleitet vom *Drosophila*-Gen *Brahma/brm*) und Yaf9, ENL, AF9, Taf14, Sas5 (YEATS)-Domänen erkannt (*6, 7*). Da diese Lesedomänen u.a. die Transkription von Onkogenen wie MYC mitbewirken, wurden sie in den letzten Jahren zunehmend Zielproteine für kleine organische Moleküle, die die Lesefunktion der Domäne und somit auch die Transkription der Onkogene inhibieren (*8, 9*).

Bromdomänen sind epigenetische Lesedomänen mit Spezifizität für Acetyl-Lysin an Histonenden der Histone 3 und 4. Im Menschen gibt es 46 Bromodomänen-enthaltende Proteine mit insgesamt 61 verschiedenen Bromodomänen. Die Domäne besteht aus ~110 Aminosäuren, deren Tertiärstruktur sich aus vier zentralen α-Helices αZ, αA, αB und αC und ihren verbindenden flexiblen Loops ZA, AB und BC zusammensetzt. Die tiefe Bindetasche wird von den Loops ZA und BC gebildet und weist typischerweise

ein Netzwerk aus konservierten Wassermolekülen sowie einen Asparaginrest auf, der eine spezifische Interaktion mit der Amidgruppe aus Acetyl-Lysin ausbildet (*6*).

YEATS-Domänen wurden kürzlich als neuartige Lesedomänen für Acetyl-Lysin und größere Modifikationen wie Crotonyl-Lysin entdeckt. Die unterschiedliche Spezifität wird durch eine flache, in zwei Richtungen offene Bindetasche ermöglicht. Auch die Tertiärstruktur unterscheidet sich von der der Bromodomänen: acht antiparallele β-Stränge bilden ein längliches Grundgerüst, das auf zwei Seiten von flexiblen Loops abgedeckt wird. Loops 1, 4 und 6 bilden auf einer dieser Seiten die Bindetasche, deren besonderes Charakteristikum eine aromatische Triade ist, die die π-π-Interaktionen mit der modifizierten Lysinseitenkette des Histons eingeht. Wasserstoffbrückenbindungen zwischen der Lysinseitenkette und Loop 6 sind neben der Stabilisierung durch das Histonrückgrat für die spezifische Bindung verantwortlich (*10*).

Spezifische Inhibitoren werden auch „chemische Sonden“ genannt. Chemische Sonden dienen dazu, die Funktion einer Proteindomäne aufzuklären, ohne dabei die Integrität des Proteins durch Mutation zu beeinflussen oder das Volllängenprotein in *knockout*-Experimenten gänzlich dem zellulären Kontext zu entziehen. Die Erkenntnisse aus Experimenten mit chemischen Sonden können so auch dazu verwendet werden, das Potential als Zielprotein für einen pharmazeutischen Wirkstoff zu bewerten. Chemische Sonden müssen deshalb im Gegensatz zu medizinischen Wirkstoffen weder strenge pharmakokinetische Voraussetzungen erfüllen noch müssen Nebenwirkungen minimiert werden. Vielmehr sind eine hohe Affinität, Spezifität und Zellgängigkeit wünschenswerte Eigenschaften (*11*).

Diese Thesis befasst sich mit der strukturbasierten Entwicklung von Inhibitoren für epigenetische Lesedomänen mit Spezifität für Acetyl-Lysin sowie der strukturellen Analyse dieser Domänen. In drei Teilen beschreibe ich, wie neue Strategien entwickelt wurden, um Inhibitoren für neue Bromodomänen zu finden, die Entwicklung eines dualen Inhibitors für eine Bromodomäne und eine Kinase sowie die Entwicklung der ersten Inhibitoren und einer chemischen Sonde für die Familie der YEATS-Domänen.

Zu Beginn dieser Promotion waren chemische Sonden für ca. 50% aller im Menschen vorkommenden Bromodomänen verfügbar; die meisten von ihnen für Vertreter der Familien II und IV (*12*). Zu ihren bekanntesten Vertretern zählt der BRD4-Inhibitor JQ1. Mit dem Ziel, Inhibitoren für Bromodomänen zu finden, für die bisher keine chemische Sonden verfügbar waren, verfolgten wir in Zusammenarbeit mit V. Myrianthopoulos einen Ansatz in drei Schritten: virtuelles Screening gefolgt von biophysikalischer Charakterisierung der Hits und anschließender Optimierung der Moleküle nach Analyse der experimentellen Daten (*13*). Im ersten Schritt wurden drei virtuelle Screening-Techniken kombiniert, um

eine Sammlung aus 260000 Molekülen des NCI/DTP zu untersuchen. Für 2D- und 3D-Ähnlichkeitsanalysen wurden die Moleküle der Sammlung mit JQ1 verglichen, da diese Analyse vor Beginn der Promotion durchgeführt wurde und diese chemische Sonde der einzige potente Bromodomänen-Inhibitor war. Um JQ1-ähnliche Pharmakophore zu vermeiden, wurden alle Benzodiazepine von der Analyse ausgeschlossen. Die 5000 ähnlichsten Moleküle aus der 2D-Ähnlichkeitsanalyse wurden weiterhin in einer Dockingstudie auf ihre potentielle Interaktion mit der Acetyl-Lysin-Bindetasche bewertet. Die besten 40 Moleküle aus den drei Methoden, 2D/3D/Docking wurden biophysikalischen Analysen unterzogen. *Differential Scanning Fluorimetry* (DSF), das auf veränderten Proteinschmelzpunkten bei Zugabe von Inhibitoren basiert, und isothermale Titrationskalorimetrie (ITC) dienten zur Charakterisierung der Moleküle. Aus dieser Auswahl konnte ein Hit für die fünfte Bromodomäne von Polybromo 1 (PB1(5)) mit einer Affinität von 11.5 µM identifiziert werden, der wenige *off target*-Effekte aufwies und aufgrund seiner chemischen Struktur Modifikationen ermöglichte. Der Hit würde als Grundlage für eine rechnerische Analyse der in der Bindetasche befindlichen Wassermoleküle verwendet. Interessanterweise wurde ein instabiles Netzwerk aus Wassermolekülen in der Peripherie des ZA-Kanals entdeckt, deren Platz durch Alkylverlängerungen im Molekül eingenommen werden sollte – von Ethyl bis Butyl. In der darauffolgenden biophysikalischen Analyse, die in unserem Labor durchgeführt wurde, konnte in der Tat gezeigt werden, dass die Modifikationen einen Affinitätsanstieg von 11.5 µM auf 3.3 µM bewirkte (*13*).

Der zweite Teil dieser Dissertation befasst sich mit der Entwicklung von dualen Bromodomänen-Kinase-Inhibitoren. Ciceri *et. al* zeigten im Jahr 2014, dass einige der publizierten Kinase-Inhibitoren unabsichtlich Bromodomänen der Familie II inhibierten, zu denen auch BRD4 gehört. Zu ihnen zählten TG-101348 und BI-2536, welche beide als Polo-like kinase 1 (PLK1)-Inhibitoren entwickelt wurden (*14*). Ziel war es, das Grundgerüst von BI-2536 zu nutzen, um einen dualen BRD4-ALK^{F1174L}-Inhibitor zu entwickeln. Die F1774L-Mutation tritt in schweren Neuroblastomfällen auf und führt zu erhöhten MYCN-Leveln, die durch den *Super Elongation Complex* auch über BRD4 reguliert werden (*15, 16*). Daher wurden vier Stellen in BI-2536 systematisch verändert, um eine Verringerung der Affinität zu PLK1 und eine Erhöhung der ALK^{F1174L}-Affinität zu erreichen. Zu diesen vier Modifikationsstellen zählte auch die Methoxygruppe im Aromaten neben dem Dihydropteridinon, dem eine zentrale Rolle für die PLK1-Selektivität zugesprochen wurde. Nach ausführlicher Betrachtung der Struktur-Aktivitätsbeziehung wurde Molekül B3 als dualer BRD4-ALK^{F1174L}-Inhibitor identifiziert. Seine Affinität zu BRD4 betrug 23 nM gegenüber ALK^{F1174L} und 44 nM gegenüber BRD4. Die Messung mitotischer Marker in HEK293T-Zellen zeigte indes die Eliminierung der PLK-Aktivität bis 10 µM Inhibitorkonzentration (*16*).

Der dritte Fokus dieser Dissertation waren YEATS-Domänen. Ko-Kristallstrukturen und Peptidassays mit Histon 3-Peptiden zeigten, dass die nach hinten offene Bindetasche verlängerte Lysinmodifikationen wie Crotonylfunktionen durch Interaktion mit der aromatischen Triade akzeptierte (*17–20*). Außerdem wiesen die *knockout*-Studien von Erb *et. al* darauf hin, dass die YEATS-Domäne von ENL maßgeblich an der Entstehung von akuter myeloischer Leukämie (AML) beteiligt ist (*21*). Diese Ergebnisse gaben Anlass, diese YEATS-Domäne für weitere strukturelle Untersuchungen und als potentielles Target mithilfe chemischer Sonden zu untersuchen.

Nachdem die stabile Expression des Proteins gewährleistet war, bot sich in unserem Labor die Hochdurchsatzuntersuchung mehrerer organischer Moleküle mithilfe des DSF-Assays an. Jedoch wiesen alle vier humanen YEATS-Domänen hohe Schmelzpunkte auf, so dass stabilisierende Effekte durch Inhibitoren erst bei Inhibitorkonzentrationen sichtbar wurden, die mit der Farbstofffluoreszenz interferierten. Daher fiel die Entscheidung für dieses Projekt auf einen strukturbasierten Ansatz mit Röntgenkristallographie als Hauptmethode (*22*). Zunächst wurde eleven-nineteen-leukemia protein (ENL) YEATS ohne Liganden kristallisiert, um einerseits zu verstehen, welche Konformationsänderungen mit der Histon- bzw. Ligandenbindung einhergehen und andererseits um eine kristallografische Plattform zur Inkubation von ligandenfreien (bzw. Apo-) Proteinkristallen mit Inhibitorlösungen zu etablieren. Die Apo-Struktur von ENL YEATS wurde mit einer Auflösung von 1.8 Å gelöst und gab Aufschluss über die Beschaffenheit der Bindetasche. Die Seitenkette von Y78 - ein Rest der aromatischen Triade - zeigte deutliche Flexibilität mit einer "in"-Position, die mit Beobachtungen aus publizierten histongebundenen Strukturen übereinstimmte, und einer "out"-Position, die die Bindetasche schneisenförmig öffnete. Außerdem wurden zwei konservierte Wassermoleküle identifiziert, die durch Wasserstoffbrücken mit dem Rückgrat von Loop 6 stabilisiert wurden. Als nächstes versuchten wir, Acetyl-Lysin als solches ohne Peptidrückgrat mit den Apo-Kristallen zu inkubieren. Es zeigte sich ausgeprägte Elektronendichte an der Stelle eines der konservierten Wassermoleküle, die mit der Stelle der Amidgruppe in der histongebundenen Struktur übereinstimmte. Obwohl die Elektronendichte nicht das gesamte Acetyl-Lysin beinhaltete, schlussfolgerten wir, dass die Aminosäure gebunden hatte, dass jedoch das Cα-seitige Ende der Aminosäure durch das fehlende Peptidrückgrat flexibel sein musste und folglich zu keiner definierten Elektronendichte führte. Darauf aufbauend wählten wir für die Kristallinkubationsstudie Moleküle mit Amidbindung flankiert von einem oder zwei aromatischen Resten aus, die potentiell mit der aromatischen Triade in der hinteren Bindetasche agieren konnten. Aus 19 inkubierten Molekülösungen gingen zehn kristallografische Datensätze mit ENL YEATS hervor, von denen zwei genügend Elektronendichte aufwiesen, um die

Bindekonformation der Inhibitoren zu modellieren – C12 und C19. Beide enthielten Aromaten sowohl auf der Seite der Carbonylgruppe des Amids (R1) als auch auf der Aminseite (R2). C12 und C19 zeigten verglichen mit der publizierten histongebundenen Struktur eine umgekehrte Orientierung der Amidgruppe, so dass R1 in Richtung H56 zur Vorderseite der Tasche zeigte und sich parallel zu H56 ausrichtete. Während Y78 im Fall von C12 eine „in"-Position annahm und R2 π- π-Wechselwirkungen mit der aromatischen Triade einging, wies Y78 bei C19 eine „out"-Position auf und R2 ragte mit seinem chlorsubstituierten Aromaten aus der länglichen Tasche hinaus. C12 und C19 band ENL YEATS mit zweistellig-mikromolarer Affinität. Mit einer Ko-Kristallstruktur von ENL YEATS mit C20, einem Inhibitor aus einem Screen unserer Kollaborateure aus Oxford (*23*), zeigten wir außerdem, dass man die Affinität unter 1 µM erhöhen konnte, in dem man R2 durch Benzimidazol-Piperidin ersetzt, das weitere Wasserstoffbrücken mit Loop 6 und der Seitenkette von E75 einging. Die erste chemische Sonde für eine YEATS-Domäne, SGC-iMLLT, ein Derivat von C20, wies einen analogen Bindemodus auf und erreichte Affinitäten von jeweils 129 nM und 77 nM für ENL YEATS und das verwandte AF9 YEATS. Intrazelluläre Interaktion zwischen SGC-iMLLT wurde mithilfe von *Fluorescence Recovery after Photobleaching* (FRAP) eines zellulären *Thermal Shift Assays* (CETSA) und für AF9 mithilfe des NanoBRET-Assays bewiesen. Schließlich wurde die Transkription MYC, Dendrin und CD86 in MV4;11-Zellen verändert, was die Relevanz von ENL/AF9 im AML-Kontext betont (*24*).

Das Ziel dieser Thesis war es, die rationale Entwicklung von Inhibitoren für epigenetische Lesedomänen mit strukturbiochemischen Arbeiten zu unterstützen. Dieses Ziel wurde in drei Teilaspekten erfolgreich erreicht. Nach dem großen Erfolg von JQ1 als chemische Sonde für BRD4 konnten wir in Zusammenarbeit mit V. Myrianthopoulos *et. al* ein Netzwerk aus unstrukturierten Wassermolekülen nutzen, um ein im virtuellen Screening gefundenes Hit-Molekül dreifach stärker an PB1(5)/SMARCA2/4 binden zu lassen (*13*). In Kollaboration mit E. Watts *et. al* vom ICR in London modifizierten wir den PLK1-BRD4-Inhibitor BI-2536 an vier Stellen, so dass seine Spezifität von PLK1 zu ALK^{F1174L} geändert wurde, ohne BRD4-Affinität einzubüßen. Dieser Inhibitor verspricht ein wichtiges Mittel zu werden, um schwierige Neuroblastomfälle mit dieser Mutation zu behandeln (*16*). Schließlich hat diese Dissertation maßgeblich zum strukturellen Verständnis von YEATS-Domänen und zur Entwicklung der ersten chemischen Sonde für diese Proteinfamilie beigetragen. Dies beinhaltet die erste Apo-Struktur einer humanen YEATS-Domäne und die ersten Ko-Kristallstrukturen zwischen einer YEATS-Domäne und kleinen organischen Molekülen (*22, 24*). Diese Ergebnisse stellen die Grundlage für weitere Fortschritte für die Entwicklung von chemischen Sonden für andere humane YEATS-Domänen dar und lassen auf seine Nutzung in weiteren zellulären Untersuchungen zum tieferen Verständnis von AML hoffen.

Zusammenfassung (kurz)

Epigenetische Mechanismen regulieren, wie die Information der DNA in verschiedene Phänotypen umgesetzt wird. DNA-Methylierungen und post-translationale Modifikationen an Histonenden nehmen hierbei eine wichtige Rolle ein und bilden gemeinsam die sog. epigenetische Landschaft. Diese Modifikationen regulieren die Genaktivität und dienen als Plattformen für transkriptionsrelevante Bindungspartner (*2*). Sogenannte “Schreiber” sind zuständig für das Hinzufügen der Modifikationen, “Löscher“ entfernen sie und „Leser“ bzw. Lesedomänen erkennen sie spezifisch und vermitteln den Kontakt zu anderen Proteinen. In Bezug auf Acetylierung an Histonenden, die ein Fokus dieser Dissertation sind, werden die Schreiber Histonacetyltransferasen (HAT), die Löscher Histondeacetylasen und die Lesedomänen Bromodomänen und Yaf9, ENL, AF9, Taf14, Sas5 (YEATS)-Domänen genannt (*4–7*). Eine anomale epigenetische Landschaft oder mutierte Formen von Lesedomänen können zu Krankheiten wie Krebs und Entzündungskrankheiten führen (*3*). Seit der Entdeckung der chemischen Sonde JQ1 sind Bromodomänen verstärkt Gegenstand epigenetischer und medizinischer Forschung (*9, 12*). Viele weitere chemische Sonden für Bromodomänen wurden entwickelt und einige BRD4-Wirkstoffe sind in klinische Studien eingetreten (*25*). Bromodomänen bestehen aus einer alpha-helikalen Grundstruktur und einer tiefen Bindetasche. YEATS-Domänen hingegen wurden vor einigen Jahren als epigenetische Lesedomänen mit Spezifität für Acetylierungen und größeren Acylierungen wie Crotonylierungen identifiziert. Ihre Tertiärstruktur wird von acht antiparallelen beta-Strängen gebildet und beinhaltet eine von drei Loops geformte, tunnelförmige Bindetasche (*10*).

Der erste Teil dieser Thesis behandelte Strategien für die Entwicklung neuer Bromodomänen-Inhibitoren. Anfang 2016 waren chemische Sonden für etwa die Hälfte aller humanen Bromodomänen verfügbar (*26*). In Zusammenarbeit mit V. Myrianthopoulos und E. Mikros der Universität Athen entdeckten wir ein Hit-Molekül für die fünfte Bromodomäne von Polybromo 1 (PB1(5), ein Mitglied der Subfamilie VIII, dessen Vertreter häufig in verschiedenen Krebstypen mutiert sind. Der Hit wurde mit biophysikalischen Methoden validiert und wies eine Affinität von 11.5 µM gegenüber PB1(5) auf. Durch den Inhibitor wurden vier konservierte Wassermoleküle in der Bindetasche verdrängt, was eine Besonderheit innerhalb der Bromodomänen-Inhibitoren darstellt. Drei Derivate dieses Moleküls wurden auf Basis von rechnerischen Analysen unstrukturierter Wasser in der Nähe des ZA-Kanals synthetisiert, was zu einer dreifach-besseren Verbindung in Bezug auf Affinität (3.3 µM) gegenüber PB1(5) führte (*13*).

Im zweiten Teil entwickelten wir zusammen mit E. Watts und S. Hoelder des *Institute of Cancer Research* (ICR) in London einen dualen BRD4-ALK^{F1174L} Bromodomänen-Kinase-Inhibitor. Einige akute

Fälle von Neuroblastomen mit ALK^{F1174L}-Mutation zeichnen sich durch erhöhte MYCN-Expressionslevel aus. Die alleinige Inhibition der ALK-Kinasefunktion zeigt in der Regel nur kurzfristig Wirkung. Als Teil des *Super Elongation Complex* ist BRD4 am Elongationsprozess beteiligt und reguliert somit auch die Transkription von Onkogenen wie MYCN. Durch systematisches Modifizieren des PLK1-BRD4-Inhibitors BI-2536 an vier Stellen entwickelten wir einen potenten BRD4-ALK^{F1174L} Inhibitor mit 23 nM Affinität gegenüber ALK^{F1174L} und 44 nM Affinität gegenüber BRD4(1). Der Inhibitor löste keine PLK1-Aktivität in in HEK293T-Zellen bis 10 µM aus und intrazelluläre Interaktion mit ALK und BRD4 wurde in einem NanoBRET-Experiment gezeigt (*16*).

Der dritte Teil behandelte YEATS-Domänen und die Entwicklung von Inhibitoren für diese neue Familie epigenetischer Lesedomänen. Eleven-nineteen-leukemia protein (ENL) und ALL1-fused gene from chromosome 9 protein (AF9) sind wie BRD4 Teil des *Super Elongation Complex* und häufig Fusionspartner von mixed lineage leukemia protein (MLL) in akuter myeloischer Leukämie (*19, 21*). Diese Dissertation enthält die erste Apo-Struktur (ligandenfreie Struktur) von ENL YEATS. Sie zeigt Flexibilität der Y78-Seitenkette in der für dieses Protein typischen aromatischen Triade und zwei konservierte Wassermoleküle. Ko-Kristallstrukturen mit kleinen organischen Molekülen definierten weiterhin Voraussetzungen für ENL YEATS-Inhibitoren. Die entdeckten Inhibitoren enthielten eine zentrale Amidgruppe, die eines der beiden konservierten Wassermoleküle verdrängte. Die Amidbindung bildete eine beta-Faltblatt-artige Interaktion mit dem Rückgrat von Loop 6 und der S58-Seitenkette aus und war flankiert von zwei Aromaten, von denen einer mit der H56-Seitenkette in der vorderen Bindetasche und einer mit der aromatischen Triade in der hinteren Bindetasche interagierte (*22*). Die mit M. Moustakim *et. al* (Oxford Universität) publizierte chemische Sonde für ENL/AF9 war die erste Sonde für eine YEATS-Domäne. Unsere Ko-Kristallstrukur erklärte, dass die Affinität in den niedrigen nanomolaren Bereich verbessert werden konnte (ENL: 129 nM, AF9: 77 nM), in dem weitere Wasserstoffbrücken zwischen dem hinteren 2-Methyl-Pyrrolidin-Benzimidazol und dem Loop 6-Rückgrat sowie der E75-Seitenkette ermöglicht wurden. Die Sonde veränderte die Transkriptionslevel von MYC, Dendrin und CD86 in MV4;11-Zellen, was die Relevanz von ENL/AF9 im AML-Kontext betont (*24*).

Diese Dissertation trug wesentlich zu drei aktuellen Themen im Bereich chemischer Sonden für epigenetische Lesedomänen bei: 1) die Entwicklung von Inhibitoren für die Bromodomänen-Subfamilie VIII auf Basis struktureller Besonderheiten, 2) duale Inhibitoren, die synergistische Effekte verschiedener Proteinfamilien nutzen und 3) strukturelle Eigenschaften der ENL YEATS-Domäne, ergänzt durch mehrere Ko-Kristallstrukturen und schließlich die Entwicklung der ersten chemischen Sonde für eine YEATS-Domäne.

Summary

Epigenetic mechanisms largely influence how genetic information on DNA level is translated into different phenotypes. DNA methylations and histone post-translational modifications make up what is referred to as "epigenetic landscape", an interconnected pattern that regulates access to genes and serves as platform for specific binding partners (*2*). The epigenetic landscape is maintained by "writers", which add the modifications, "erasers", which delete the modifications and "readers" which specifically bind modifications and mediate their location to other proteins connected to transcription. In the context of acetylations, which are the focus of this thesis, the writers are called histone acetyl transferases (HATs), the erasers are called histone deacetylases (HDACs) and the readers comprise Bromodomains (BRDs) as well as Yaf9, ENL, AF9, Taf14, Sas5 (YEATS) domains (*4*–*7*). An aberrant epigenetic landscape and mutated forms of epigenetic readers can lead to diseases including cancer and inflammatory diseases, making epigenetic reader domains attractive drug targets (*3*). Not only since the discovery of JQ1, the first chemical probe for a BRD (BRD4), BRDs have been heavily investigated (*9, 12*). Many more chemical probes have been released and BRD4-targeting drugs have entered clinical trials (*25*). They feature an alpha-helical tertiary structure and a groove-like binding pocket. YEATS domains, however, only recently emerged as epigenetic reader domains with specificity for acetylations as well as larger lysine modifications like crotonylation. Their basic structure consists of eight antiparallel beta-strands and a cave-like pocket, which is open on two sides, being formed by three loops protruding from the beta sheets (*10*).

The first part of the thesis treated new strategies to develop inhibitors for so far untargeted BRDs. At the beginning of this thesis, chemical probes were available for approx. 50% of all human BRDs with varying degrees of predicted druggability (*12*). In cooperation with V. Myrianthopoulos and the Mikros lab from the University of Athens, we combined three virtual screening approaches to discover a hit for the fifth BRD of Polybromo 1 (PB1(5)), member of family VIII, whose members are frequently mutated in many cancer types. The hit was validated by biophysical methods, showing an affinity of 11.5 µM towards PB1(5) and displacement of conserved water molecules within the binding pocket, a token not commonly detected in other BRD families. Three derivatives of the hit were synthesised on basis of a computational analysis of unstructured water molecules in the vicinity of the ZA channel, leading to a 3-fold affinity increase up to 3.3 µM for PB1(5) (*13*).

In the second part, we developed a dual BRD4-ALK^{F1174L} BRD-kinase inhibitor in collaboration with E. Watts from the Hoelder lab in the Institute of Cancer Research (ICR), London. Severe cases of

neuroblastoma patients with this ALK mutation show elevated MYCN levels and usually the effects of sole ALK inhibition are short-termed. As part of the super elongation complex, BRD4 is responsible for release of paused Pol II at transcription elongation checkpoint control, a mechanism which also drives oncogenic transcription including MYCN. Based on BI-2536, a PLK1-BRD4 inhibitor, modifications at four positions of the molecule led to the discovery of a potent BRD4-ALK^{F1174L} inhibitor with 23 nM affinity towards ALK^{F1174L} and 44 nM affinity towards BRD4(1). PLK1 activity in HEK293T cells was eliminated up to 10 µM and target engagement towards ALK and BRD4 was shown in a NanoBRET experiment (*16*).

The third part covered YEATS domains and the development of inhibitors for this new class of epigenetic readers. Eleven-nineteen-leukemia protein (ENL) and ALL1-fused gene from chromosome 9 protein (AF9) are also part of the super elongation complex and are common fusion partners of mixed lineage leukemia protein (MLL) in acute myeloid leukemia (AML) (*19, 21*). In this thesis, the first apo crystal structure of ENL YEATS revealed an inherent flexibility of the Y78 side chain in the aromatic triad and two conserved water molecules. Soaking experiments led to the first co-crystal structures between a YEATS domain and small molecule inhibitors and defined prerequisites for ENL YEATS inhibitor scaffolds. The discovered inhibitory fragments had a central amide bond in common, which replaced one of the two conserved water molecules to form beta-sheet-like hydrogen bonds between the loop 6 backbone and the S58 side chain. The amide bond was flanked by two aromatic moieties, of which one stacks with H56 in the front pocket and the other interacts with the aromatic triad in the rear pocket (*22*). The development of the first chemical probe for ENL/AF9, SGC-iMLLT, show that the affinity is increased to low nanomolar levels if the rear flanking aromatic moiety forms additional hydrogen bonds with loop 6 and the side chain of E75. In case of the probe, this is achieved with a 2-methyl-pyrrolidine-benzimidazole moiety. The probe binds with high affinity to ENL (129 nM) and AF9 (77 nM) and shows no significant affinity towards other human YEATS domains or BRDs. Target engagement was shown by fluorescence recovery after photobleaching (FRAP), cellular thermal shift assay (CETSA) and in case of AF9 also with NanoBRET. The probe changed the expression of three AML-related genes (MYC, dendrin and CD86) in MV4;11 cells, encouraging application of this probe in more AML cell lines (*24*).

This thesis contributed to three current topics in the field of chemical probes for epigenetic reader domains: 1) exploiting structural features of untargeted BRD families, 2) the development of dual inhibitors, which make use of synergistic effects in different target families, and most importantly 3) structural features of the ENL YEATS binding pocket, the first small molecule inhibitors accompanied by the first apo-structure of a human YEATS domain as well as the first co-crystal structures with small molecules.

Introduction

Epigenetics

What does the term "Epigenetics" mean?

When Francis Crick and James Watson unravelled the three-dimensional structure of deoxyribonucleic acid (DNA) in the year 1953 (*27*), scientists hoped to finally understand the underlying process of cell physiology and the molecular mechanism that compose human phenotypes. Watson and Crick's work revealed not only a molecule structure, but surprised science by relating the construction of living organism to a mere combination of four DNA bases. In the following years, the process of gene transcription to messenger ribonucleic acid (mRNA) and subsequent translation to proteins by the ribosome have been major discoveries and formed for a long time a scientific dogma about how phenotypes are formed (*28*). The discoveries also prompted the equally momentous Human Genome Project, which was aimed at sequencing the entire human genome, an effort made over 13 years by a vast number of public institutions and private enterprises (*29*). Today, we approach a new era which allows the sequencing of individual genomes with the help of next generation sequencing methods. Certainly, the efforts to improve sequencing genome sequencing methods is largely driven by the realisation that diseases like cancer need treatment that precisely suits the given case (*30*).

However, although the identification of genes and their potential influence on certain diseases represented a huge step in our understanding of molecular mechanisms leading to illnesses, the genetic information alone is not sufficient to understand the function and role of genes and of proteins in their respective pathways. To date, 17,000 of the 19,000 human genes are still underinvestigated, partially because research strategy is derived from model organisms like *Drosophila melanogaster (31)*. Also, the discovery of non-coding RNA revealed that only a fraction of genes actually translates to proteins, which – despite its historical importance – exposes the gene-to-protein dogma as an oversimplification of the matter in hand (*32*). Furthermore, while it is indeed important to investigate the function of genes, it is at least as important to learn about how they are regulated. The term "Epigenetics" literally means "above genetics", thus describing mechanisms that control the activity of genes (*33*). The necessity to control gene activity is evident because of the fact that the whole genome is present in most cells of the human organism, yet the cells are highly differentiated to fulfil a distinct purpose. A prominent example for epigenetic control is the silencing of one of the female X chromosomes (XCI), a process occurring during early embryogenesis (*34*). Throwing light on the function of proteins is a process inevitable to the

understanding of function and disease and is one of the four major fields of research in the Structural Genomics Consortium (SGC). It is a not-for-profit organisation, consisting of public institutions and pharmaceutical companies aimed to fill this knowledge gap through structural information and biochemical elucidation. The collection of protein structures from metabolic enzymes, receptor signalling proteins, human protein kinases and epigenetic proteins are four major areas of contribution, through which targeting of these proteins with small molecule inhibitors has been greatly facilitated (*35, 12*).

In general, epigenetic regulation refers to modulation of gene accessibility and therefore gene activity. DNA is coiled around an octamer of histone proteins – two H2A, H2B, H3 and H4 each – into nucleosomes. Positively charged histone tails on the N-terminus mediate the interaction with the negatively charged DNA backbone. In a rather open form suitable for transcription of a gene, these nucleosomes are interconnected by DNA strands to form euchromatin, a structure of approx. 10 nm in diameter looking like “beads on a string”. In structurally restricted areas of the genome, the nucleosomes wrap around histone H1 proteins, leading to a 30 nm wide coiled structure called heterochromatin. Epigenetic regulation comprises all chemical modifications on chromatin (either DNA or post-translational modifications (PTMs) on histones) or structural influences like chromatin remodelling inter-/intrachromosomal DNA-DNA interactions. More specifically, DNA can be methylated and histone tails are known to be methylated, acetylated (and various acylation marks), phosphorylated, sumoylated and ubiquitinated (*1*). The focus of this thesis is histone tail acetylation marks, so more detailed descriptions will treat its key players and implications.

The sum of epigenetic modifications is often referred to as “epigenetic landscape”, a metaphor that was first used by Conrad Hal Waddington in 1957. The term describes how the fate of a cell is determined by addition of distinct epigenetic modifications (Figure 1). The cell is represented by a ball that reaches valleys in the diagram standing for different cell fates. On the way, certain variables Ψ (certain sets of epigenetic modifications) determine the potential Φ between the valleys (*2*).

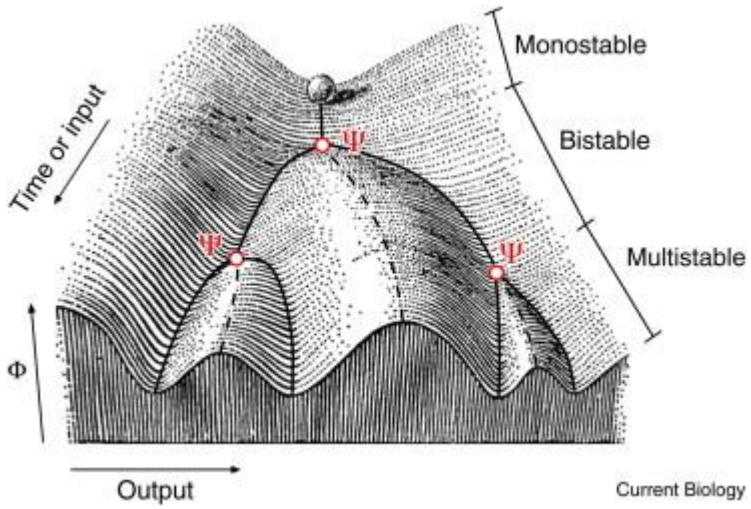


Figure 1 The "epigenetic landscape". Time and the addition of epigenetic modifications lead to certain cell fates, represented by valleys in the plot. The cell outcomes are influenced by variables Ψ separated by potentials Φ (adapted from (*36*) and (*2*), Copyright (2012), with permission from Elsevier).

The potential surface is defined by the system of genes, RNAs and proteins that influence the development of a cell into different cellular traits. However, these changes are not to be understood as program that is imprinted into the genome and is run as a cascade for development of an organism, but rather as a dynamic process that also depends on external stimuli like metabolites. In that sense, cells are able to add epigenetic modifications, act upon their effects and also remove certain modifications to create a landscape or dynamic equilibrium to fit the actual needs (*37*). Proteins that catalyse the corresponding reactions are known as epigenetic writers, readers and erasers. Histone tail lysine residues are acetylated by histone acetyl transferases (HATs), read by bromodomains, Yaf9, ENL, AF9, Taf14, Sas5 (YEATS) domains and some double plant homeodomain (PHD) finger domains and erased by histone deacetylases (HDACs) (*4, 5, 38, 10*).

Although valleys in the epigenetic landscape stand for regular, healthy differentiation states of the cell, they also comprise unbalanced states that lead to diseases. Disrupting the enzymatic function of writers and erasers as well as the recognition function of readers therefore offers opportunity to restore cells to a healthy epigenetic setup. Epigenetic writers, readers and erasers have been extensively targeted in cancer, inflammatory diseases and others (*12, 39–41*).

Important epigenetic marks and their role in disease

One of the most-studied epigenetic modifications is the methylation of cytosine bases in DNA strands. DNA methyltransferases (DNMTs) catalyse the methylation of cytosine in CpG dinucleotides, which tend to cluster in so-called CpG islands. DNMT3A and DNMT3B are responsible for *de novo* methylation, whereas DNMT1 has a maintenance role. DNA methylation is associated with silencing of genes in embryonic stem cells and includes heritable and inheritable signatures. Methylation is also involved in the aforementioned silencing of X chromosomes (*34*). Accordingly, DNMT expression levels are inversely proportional to the differentiation level of a cell. Inhibition of DNMTs by 5-aza-2'-deoxycytidine (5Aza) has been shown to activate tumour suppressor genes and is used in first-line anti-leukemic treatment (*1*).

PTMs on histone tails constitute the other part of epigenetic modifications and influence gene activity by modulating the interaction strength between DNA and histones. There are over 130 PTMs on histone proteins, including methylation, acetylation, phosphorylation, sumoylation and ubiquitination, altogether representing the so-called histone code. Due to the scope of this thesis, the focus of this paragraph is methylation and various acylations. Depending on the modification site and type of modification, activation and silencing of transcription are observed (*1*).

Histone methylation either occurs as mono-, di- or trimethylation of lysine residues or as arginine methylation. Two families catalyse the methylation in humans: Su(var)3-9, Enhancer-of-zeste and Trithorax (SET) domain-containing enzymes methylate lysine residues on histone tails (protein lysine methyltransferases (PKMTs)) and dimeric Rossmann fold protein arginine methyltransferases (PRMTs) catalyse methylation of lysine residues. Disruptor of telomeric silencing 1-like (DOT1L) is a monomeric Rossmann fold protein and transfers methyl groups to a lysine in the structured octameric histone core (*41*). DOT1L is recruited to H3K9ac sites, where it binds eleven-nineteen-leukemia protein (ENL) and trimethylates H3K79 for active transcription (*7*). Known readers of histone methylation marks are Tudor, chromo, malignant brain tumour (MBT), pro-trp-trp-pro (PWWP) domains as well as PHDs (*42*). Tripartite motif-containing protein 33 (TRIM33) contains a tandem PHD/bromodomain and binds H3K9me3 and the adjacent H3K18ac, one example showing that different histone modifications influence each other and often have to be judged in context (*43*). Histone methylation patterns are known to influence disorders and have been probed as pharmaceutical targets. For instance, inhibitors are available for proteins SET and MYND domain-containing protein 2 (SMYD2), Spindlin1 and Lysine-specific histone demethylase 1A (LSD1) (*41, 44*).

Acetyl modifications constitute another dynamic set of histone PTMs, that are added to histone tails by HATs and removed by HDACs (*4, 5*). Two processes determine the cellular effects of histone acetylation. The first process is the regulation of nucleosome assembly. From a chemical point of view, N-ε-acetylations on lysine side chains (Kac) remove the positive charge and loosen the interaction with the negatively charged DNA phosphate backbone, leading to transcription activation. In many cases, this happens in crosstalk with methylation patterns. Actively transcribed genes exhibit high levels of H3K4me3, H3K27ac, H2BK5ac and H4K20me1 in the promoter region and H3K79me1 and H4K20me1 modifications within the gene (*3*). Readers of acetylation marks, bromodomains and YEATS domains, serve as interaction modules that recruit regulatory machineries like chromatin remodelling complexes and transcription factors (*7, 6*). Although acetylation marks are found on all four types of histones, interactions between bromodomains and H3 and H4 acetylation are most prominent (*3, 8*). The second process is the influence on chromatin folding. Methylation and acetylation regulate the transition between the transcriptionally active euchromatin and the inactive heterochromatin. Euchromatin is characterised by high levels of acetylation and trimethylated H3K4, H3K36 and H3K79, whereas heterochromatin shows fewer acetylation and high levels of H3K9, H3K27 and H4K20 methylation (*3*). Disruption of the acetylation patterns are known to cause different types of cancers, inflammatory diseases and others. Bromodomains have been heavily investigated as drug targets in the past years, owing to their influence on oncogene transcription and good druggability (*12, 26*). Figure 2 shows important acetylation marks and inhibitors that intervene bromodomain binding.

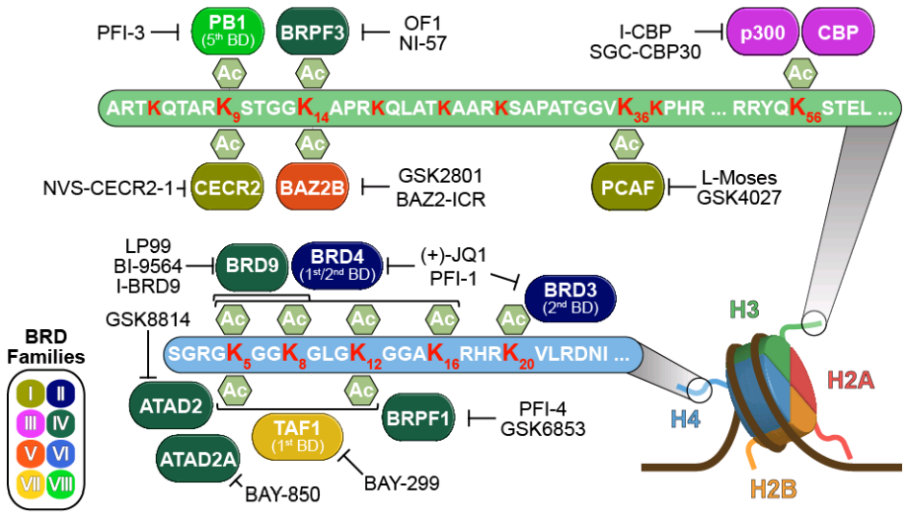


Figure 2 Bromodomain interactions with acetylation marks on H3 and H4. Bromodomain families are depicted in different colours and preference of single domains towards specific acetylation sites is shown. A selection of bromodomain antagonists is shown. (adapted from (*12*), co-author agreement obtained)

In recent years, various publications have also shed light on non-acetyl lysine acylations such as propionylation, butyrylation, crotonylation, β-hydroxybutyrylation (Bhb), succinylation and others. In case of acetylations, HATs use acetyl-coenzyme A (CoA) as substrate which is produced in metabolic processes in the cytoplasm. Accordingly, other acylations require their respective CoA-derivative, many of which are directly connected to fatty acid metabolism. Depending on their cellular concentrations, short chain acylations compete with acetylations to run genetic programs evoking a cellular (post-)metabolic response (*45*). For instance, β-hydroxybutyrylation is a histone lysine modification that is highly upregulated in all four histones upon starvation (*46*). At some starvation-specific genes, H3K9ac is completely replaced by H3K9Bhb. β-hydroxybutyrate, the substrate for this modification is abundant in cases of untreated diabetes and can be used by brain and heart as energy substitute in starvation conditions, showing how unbalanced metabolite concentrations influence the epigenetic landscape and can lead to diseased cellular states (*45, 47*). As bromodomains specifically read acetylations, their binding pocket is unsuitable for recognising larger acylations and bromodomain-specific transcription factors are less likely to be recruited. YEATS domain-containing proteins are known to have a better tolerance for longer and bulkier histone PTMs and have therefore been suggested as possible mediators to metabolism-specific transcription machineries (*45, 17, 18*). This role amongst others discussed in the YEATS-specific section of this thesis is the main rational to consider this protein family as potential target for inhibitors.

From inhibitors to drugs

Drug development process

As mentioned in the upper paragraphs, the fate of a cell is tightly regulated by its epigenetic landscape and its key regulators, epigenetic writers, readers and erasers that are actively investigated as possible drug targets. Several steps mark the process from the identification of such targets to the availability of a drug on the market (Figure 3).

The first phase during drug development is the identification and validation of targets. Possible rationales that make protein domains promising targets are for example frequent presence at oncogenic transcription sites, mutations in certain diseases or altered expression levels. There are two approaches for target validation, one being biological and one making use of so-called chemical probes (*11, 48*). The latter are subject of this thesis and are discussed in the next section. Once the target is validated as promising, hits are identified. A hit is a compound whose affinity to the target is low/mediocre but

offers possibilities for modification. The term “lead” describes compounds with improved affinity but may still need modifications to match the drug requirements. In cycles of compound perturbation and biochemical/biological assay validation, the hit is developed to a potent drug that is proved to have affinity and activity towards the target (*49*).

A pre-clinical phase determines whether the drug is efficacious in animal models and indicates possible safety issues when used in mammals. The clinical phases are performed in human test patients. In phase 1, 20-100 healthy patients help determining a dosage range over 1-2 years. Phase 2 covers a similar time range and examines the efficacy of the drug as well as well as short-term side effects in 100-500 patients with disease. The efficacy is verified in phase 3 in a higher number of patients (typically 1000-5000) and in a longer timeframe of 2-3 years to obtain more reliable information on possible long-term side-effects. Finally, the clinical data has to be submitted to the responsible regulation authorities, before it is sold on the market (*49*).

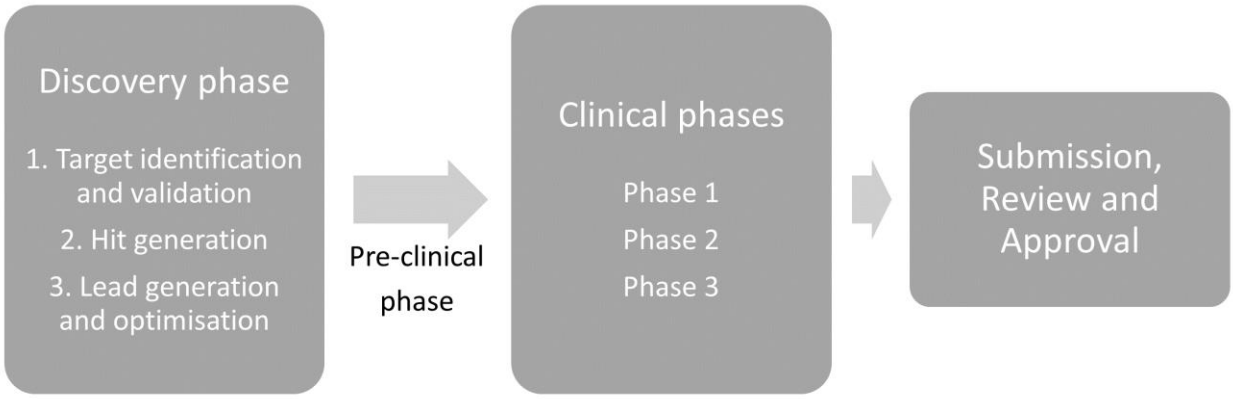


Figure 3 Steps in drug development. In the discovery phase, a target is identified and validated, and a compound is developed from a hit to a potent drug. This drug is first validated in animal models during the pre-clinical phase and further tested for doses, safety and efficacy in humans during three clinical phases. Finally, the data is submitted to authorities, before it is approved for sale on the market.

Target validation - Biological methods and chemical probes

The suitability of a protein as a target must be validated by determining its influence on the phenotype in question. Powerful biological tools have emerged in recent years, namely RNAi and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas. RNAi is a naturally occurring mechanism which uses RNA strands that are complementary to mRNA to recruit restriction enzymes. In the process, the target mRNA is degraded, and the protein of interest is not translated. Since the RNA sequence must

be known, this technique was certainly driven by the human genome project. On the other hand CRISPR/Cas is a genome editing method that allows to insert, delete or change the DNA sequence itself (*50*). Both methods are particularly helpful to investigate the protein scaffolding effects, because the whole protein is entirely withdrawn from the cellular context.

In contrast, targeting the protein of interest with a small organic molecule is especially useful to gain information about the role of a catalytic domain or distinct protein-interaction activity. Small molecules that target a protein for the purpose of target validation are called chemical probes. The effect of the probes on the target can be examined dose-dependently and their use is independent of a specific cell line. With regard to drug development, the use of a chemical probe in the target validation step also accumulates information about possible drug resistance pathways and accelerates drug development by offering possible pharmacophore candidates (*11*). Despite the numerous advantages of chemical probes, it is important to note that the mentioned biological methods answer different questions and are best used in combination.

Although the development of chemical probes may involve similar processes and methods as the development of medical drugs, their requirements differ. The most important properties of chemical probes are potency and selectivity. SGC guidelines for high quality probes are an *in vitro* affinity of <100 nM, cellular IC_{50} of <100 µM, at least 30-fold selectivity over other sequence-related proteins of the same family and profiling against sets of pharmacologically relevant off-targets and other relevant drug-targeted families (like kinases) (*11*). Other sources impose higher intra-family selectivity demands (100-fold selectivity) and add low cell toxicity levels and membrane permeability as desirable properties (*51*). Ideally, the target validation should be performed with a structurally related inactive compound as negative control and another structurally unrelated chemical probe with similar target spectrum as positive control. The latter diminishes the probability that off-target effects account for the phenotype. Medical drugs do not necessarily need to be selective, as polypharmacology may lead to desired effects, but they have stricter requirements with respect to bioavailability and certain physicochemical (molecular weight, lipophilicity) and pharmaceutic properties (stability, economic synthesis, defined crystallisation form etc.) (*11*).

Strategies in rational chemical probe design

Depending on the preliminary results and availability of structural data, either a ligand-based or a structure-based approach can be used to develop inhibitors. Either knowledge about ligands/substrates

or structural information of the binding pocket is necessary. If ligands are unknown, a close look at interaction partners and the biological function of the protein are likely to serve as starting points. If the location of the binding pocket within the protein domain is unknown, a combined approach of experimental and computational methods may serve to identify the binding site (*52*). As for epigenetic histone reader domains, the interaction partners are clear, since the binding pockets must interact with peptide backbones and must accommodate either modified lysine or arginine side chains.

Ligand-based drug discovery (LBDD) is a strategy commonly used when no three-dimensional structure of the target is available. A combination of computational and experimental data is used to establish a quantitative structure-activity relationship (QSAR) and to finally optimise the compound. It requires a set of ligands or substrates that are known to bind the protein. They should feature a combination of structural similarities and different physicochemical properties that lead to varying degrees of target activity. Furthermore, they should show to the desired biological activity. In the first step, a set of ligands is identified, and experimental data is acquired. Then, molecular descriptors of the molecules, that are associated with structural and physicochemical properties, are defined. These descriptors are correlated to the experimental data to form hypotheses which specific properties lead to changes in activity. After statistical validation of the model, the QSAR information is used to design new series of ligands that are synthesised and experimentally validated with respect to their biological effects (*53*).

The statistical validation in LBDD and a certain degree of compound diversity is necessary for optimisation and the approach may not be successful targeting newly discovered binding pockets. Structure-based drug discovery (SBDD) aims at situations, where the biological targets promise to be interesting, but none or few ligands are known. Its basis is the acquisition of a three-dimensional structure of the binding pocket, in most cases by X-ray crystallography (*54*), either in complex with ligands or as empty binding pocket (apo). A systematic crystallisation approach with small volume conditions, which accelerates the crystallisation process, allows reducing the number of trials and accelerating decisions for subsequent experimental cycles. The structural information of the binding pockets is complemented by screening either fragments or larger compounds. Fragments promise to be structurally unbiased but can be challenging to identify due to their low affinity. Possible methods for compound screening are nuclear magnetic resonance (NMR), differential scanning fluorimetry (DSF), surface plasmon resonance (SPR) or – if the apo crystallisation of the domain is established – X-ray crystallography (*54*). Depending on the flexibility of the pocket, molecular docking can supplement the knowledge about the ligand binding mode, but careful measures have to be taken to reduce bias (*55*). In X-ray crystallography-based SBDD, chemical probes are optimised in cycles of acquiring co-crystal

structures and biophysical properties like affinity, helping to understand which interactions are to be optimised, and chemical synthesis of potentially more affine compounds.

Dual inhibitors

Typically, chemical probes investigate the role of one target in context with a certain phenotype. In some diseases, the phenotype can be pinpointed to a single target and treatment strategies can be developed. In other cases, inhibition of targets takes effect, but is rapidly compensated by other biochemical pathways and nullified to the initial phenotype. Due to the interconnectivity of signalling pathways and crosstalk between epigenetic modifications and its key players, this is also possible when initial data suggests a single responsible gene/protein.

In a combination therapy, two drugs are applied to evoke a longer-lasting effect. Other signalling pathways are less likely to compensate for the effect of one drug and mutations building up against both drugs at the same time are far less likely to occur. Effects of inhibitors for different proteins or even protein classes often prove to be synergistic, meaning that they enhance their effects to more than their sum. But the development of such a synergistic approach is lengthy and expensive, because they must go through trials separately, before they can be tested in combination. Also, adverse effects occurring in addition are difficult to predict and combination therapies prove unsuitable for some patients.

Combining the desired drug effects into one molecule is pharmaceutically very challenging, but the approach significantly decreases the trial efforts and is promising to have a better impact on a patient's disease (*56*). Some drugs have been shown to have some degree of polypharmacology. For instance, imatinib is designed to inhibit the tyrosine kinase domain of the Abelson proto-oncogene (abl) in a fusion protein that is present in patients with a subtype of chronic myelogenous leukemia, but also inhibits c-kit and the platelet-derived growth factor receptor (PDGF-R). Crizotinib is used in treatment of non-small cell lung cancer (NSCLC) and inhibits the kinase domain of the anaplastic lymphoma kinase (ALK), but also potently binds c-ros oncogene 1 (ROS1) and mesenchymal-epithelial transition factor (c-MET) kinase domain (*57*). In 2014, Ciceri *et. al.* showed that two potent polo-like kinase 1 (PLK1) inhibitors, BI-2536 and TG-101348, unintendedly also inhibit the epigenetic reader domain of bromodomain-containing protein 4 (BRD4) with nanomolar affinity, indicating why these inhibitors are less prone to adaptive resistance. Also, they described the PLK1 binding mode and postulated that target specificity can be modulated rationally through modification of the inhibitors (*14*).

Epigenetic reader domains

Bromodomains

Structure and molecular recognition

Bromodomains (BRDs) specifically recognise acetyl-lysine sidechains on histone tails. There are 61 representatives in 46 diverse proteins in human. The full length proteins are responsible for recruiting transcription factors and chromatin remodelling complexes and therefore act as "interpreters" of the histone code (*6*). The name "bromodomain" derives from the *Drosophila* gene *Brahma/brm*, whose investigation lead to the discovery of the domain (*58*).

BRDs are ~110 amino acid long and are composed four core alpha helices, named αZ, αA, αB and αC, which are interconnected by three flexible loops, ZA, AB and BC (Figure 4A). The canonical binding pocket is formed between ZA and BC loop, interacting with the histone peptide backbone and accommodating acetyl-lysine in a groove-shaped pocket. The key interaction between the acetyl group and the protein is mediated by an asparagine side chain (N140 in BRD4) and a network of water molecules, mediated by a tyrosine hydroxyl group, which stabilises the interaction. There are five conserved water molecules in the binding pocket (Figure 4B and 4C). A so-called gatekeeper residue (I146 in BRD4) determines the narrowness of the ZA channel and is one attribute that distinguishes structural differences throughout BRD families (*59*). Most of acetyl-lysines recognised by BRDs are located within histones 3 and 4, where they are associated of enabling transcription by recruiting transcription (*8*). Some reports also show that adjacent lysine acetylations and methylations lead to binding of tandem BRD-PHD domains. Others publications report the binding of two acetylation marks by a single BRD in a cooperative manner (*6*)(*60*).

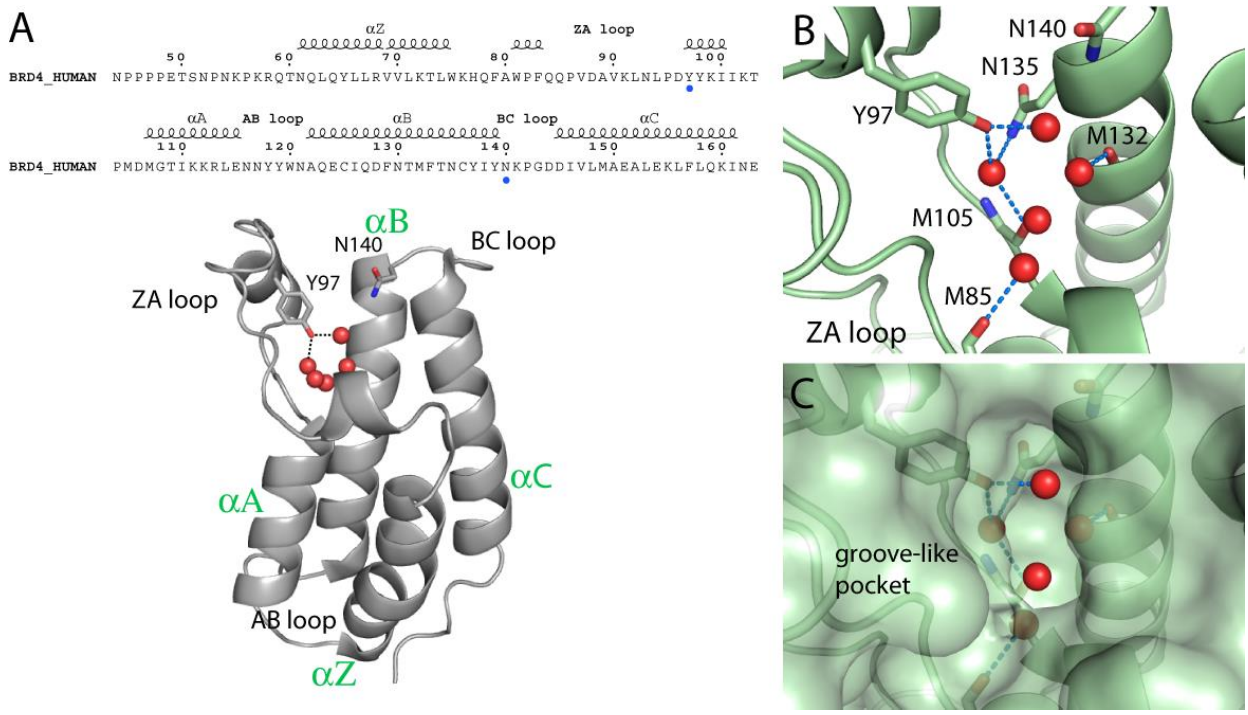


Figure 4 Composition of the BRD binding pocket **A** Overall sequence and tertiary structure of BRD4 (PDB ID: 2OSS) (*6*). The core domain is constituted by four helices and a binding pocket formed by loops ZA and BC. Blue dots represent the position of highlighted residues N140 and Y97 in the protein sequence. Partially illustrated with ESPript 3 (*61*) **B** Close-up of the BRD4 binding pocket. Five water molecules are conserved in the empty binding pocket and the ligand interaction is mediated by conserved asparagine and tyrosine side chains. **C** Surface representation of BRD4 shows a groove-like binding pocket that can accommodate acetyl-lysine.

BRD families, their involvement in diseases and available inhibitors

BRDs group into eight families based on structure/sequence similarity and have been extensively targeted by chemical probes. Figure 5 gives an overview over the bromodomain families and representative chemical probes. Proteins that have been targeted by at least one inhibitor are marked red. For all families except family V, containing the promyelocytic leukemia (PML)-SP100 nuclear bodies, at least one potent chemical probe is available (*12*).

Family I comprises the HATs P300/CBP-associated factor (PCAF) and general control nonderepressible 5 (GCN5) as well as cat eye syndrome critical region protein 2 (CECR2) and fetal Alz-50 clone 1 protein (FALZ). This family is characterised by a large tyrosine gatekeeper residue and an acidic glutamate residue in the ZA loop (compared to leucine in BRD4), usable as means for gaining specificity (*59*). The first chemical probe targeting this family was L-Moses, a dual inhibitor for PCAF/GCN5. It shows 126 nM affinity towards PCAF and 500 nM towards GCN5 (determined by ITC) and target engagement was shown via disruption of the PCAF-BRD/H3.3 interaction in HEK293 cells (*62*). Later,

GSK4027 has been published with similar dual inhibition abilities, but alternative chemical scaffold and increased affinities of 1.4 nM for both domains (determined by BROMOscan®), enabling comparative experiments to understand the role of PCAF and GCN5 in disease (*63*). The *CECR2* gene is a candidate for causing the cat eye syndrome, a chromosomal disorder leading to preauricular anomalies, anal atresia, and iris coloboma (*64*). NVS-CECR2-1 is potent CECR2-BRD inhibitor with a K_D of 80 nM (determined by ITC) and very good selectivity among BRDs and other protein families like kinases, proteases and receptors (*44*). An alternative chemical probe, GNE-886, has recently been published with improved affinity of 42 nM, but some off-target activity towards BRD9, BRD7 and transcription initiation factor TFIID subunit 1 (TAF1)/TAF1L (determined by BROMOscan®) (*65*).

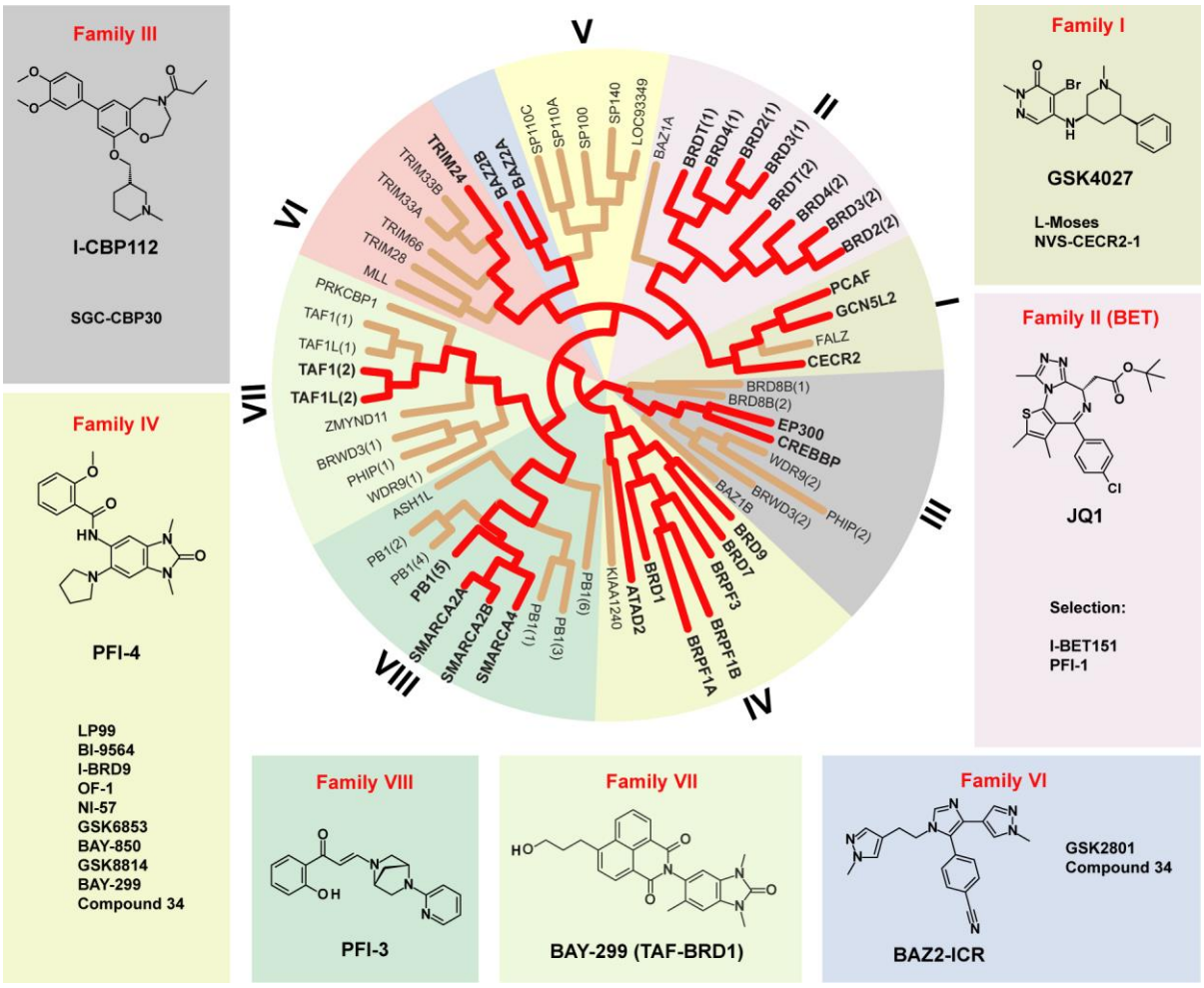


Figure 5 Bromodomain families I-VIII. Reader domains with published selective inhibitors are marked red. Boxes give overview over available inhibitors and a molecular structure for one representative.

Family II are the bromo- and extra-terminal (BET) domains. The name is derived from the fact that each of the family members features an N-terminal BRD and an extra BRD. Their tertiary structure features a long ZA loop and three stacked residues at the entrance of the pocket, called the WPF shelf (*66*) and the gatekeeper is the medium-sized and hydrophobic I146 (*59*). Due to their important role in transcription control, known phenotypical influence and excellent druggability, this is the most studied and targeted BRD family (*12*)(*26*). For its most prominent representative BRD4, around 300 structures of the protein itself and in complex with ligands are published to date (*67*). BRD4 regulates the transcription of oncogenes as part of the super elongation complex. RNA polymerase II (Pol II) is paused 30-40 nucleotides downstream of the transcription start site and released for active transcription after being phosphorylated in the carboxyterminal domain by the cyclin-dependent kinase 9 (CDK9)-subunit of the positive elongation factor b (P-TEFb). P-TEFb itself is brought into close proximity with Pol II by interaction with BRD4 and regulated in activity by COUP-TF-interacting protein 2 (CTIP2) in complex with hexamethylene bis-acetamide-inducible protein 1 (HEXIM1)/7SK RNA (Figure 6) (*15*). The pausing of the super elongation complex is frequently suppressed by mixed-lineage leukemia (MLL) fusion proteins due to genetic translocations, leading to childhood haematological malignancies (*68*). The downregulation of MYC has been shown with the first two chemical probes targeting the BET domains, JQ1 and iBET, undoubtedly showing its relevance in progression of cancer (*9*).

BRD4 has also been shown to interact with acetylated lysine residues on proteins than histones, regulating nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB)-dependent signalling, viral and metabolic diseases (*69*). An interesting perspective on relevance of BET inhibition has been proposed in context with bromodomain testis-specific protein (BRDT) as possible target for male contraception (*70*). However, despite their relevance and many synthetic efforts, selective inhibition of one BET family member has yet to be shown. Now, many publications focus on improved oral availability and pharmacokinetics and several BRD4 inhibitors are in clinical trials (*25*).

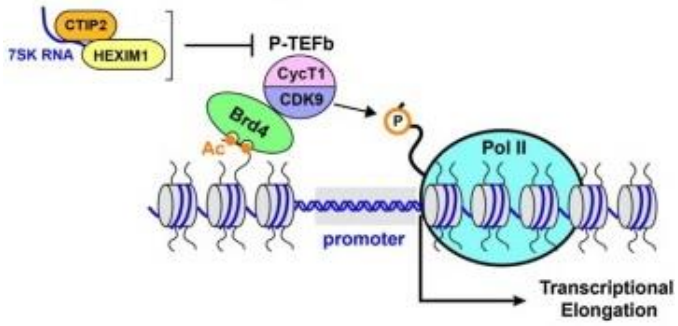


Figure 6 BRD4 and its role in transcription elongation. BRD4 localises P-TEFb at promotor sites, where it phosphorylates Pol-II through the CDK9 subunit to initiate transcription elongation. P-TEFb is inhibited by a complex consisting of CTIP2, 7SK RNA and HEXIM1. (adapted from (*15*), Copyright (2016), with permission from Elsevier)

Family III contains the HATs p300 and CREB-binding protein (CBP) among other BRDs. The motivation to develop chemical probes for this family is the involvement of p300/CBP in diverse cellular processes, many of which can lead to haematological malignancies, inflammation, and neuropsychiatric disorders (*71*). Unique amino acid changes in the BC loop (basic R1173 in CBP aligns with acidic D145 in BRD4) allow specific targeting and a smaller leucine residue at the tryptophan position of the WPF shelf in BET proteins accommodate bulkier moieties (*59*). SGC-CBP30 is the first potent inhibitor that was developed for this family with affinities of 21 nM for CBP and 38 nM for p300, but also off-target effects towards BET BRDs, which has to be taken into account due to the strong phenotypic influence of BET proteins (*72, 73*). I-CBP112 is an alternative probe published by the SGC, exhibiting K_Ds of 151 nM towards CBP and 167 nM towards p300, but also some BET off-target activity (*74*). This issue was addressed by inhibitors GNE-781 and GNE-207, both being very potent CBP binders and showing excellent selectivity towards BET BRDs (*75, 76*).

Family IV BRDs take part in HAT scaffolding with bromodomain and PHD finger containing proteins 1-3 (BRPF1-3) and in chromatin remodelling complexes with BRD7, BRD9 and ATPase family AAA domain-containing protein 2 (ATAD2) A and B (*12*). BRD7 has been shown to be a tumour suppressor, while BRD9 is either mutated, upregulated or over-expressed in several cancer types. Their part in the Switch/Sucrose Non-Fermentable (SWI/SNF) complex increases their relevance, since this chromatin remodelling complex harbours various oncogenic factors and understanding their interplay might be crucial to understand effective treatment strategies (*71*). The ZA loop of this BRD family usually adopts unique conformations. Also, tokens like the Y222 gatekeeper residue in BRD9, which restricts access to a GFF motif (corresponding to WPF in BRD4), make the requirement for different inhibitor scaffolds

evident (*59*). So far, the probes BI-9564, LP-99 and I-BRD9 are available tools for experiments on BRD7/9 (*77–79*). For ATAD2A/B, the BRD antagonists GSK8814 and BAY-850 are available, the latter being selective for isoform B through targeting of an allosteric pocket (*80, 81*). OF-1 and NI-57 are two available pan-BRPF inhibitors (*82*) and GSK6853 and PFI-4 are BRPF1B-selective (*83, 82*). Furthermore, two dual inhibitors, which enable the study of family IV in connection with other BRD families, have been reported. BAY-299 is an inhibitor for BRPF2 and TAF1(2) and is the only probe with selectivity towards family VII (*84*). Compound 34, on the other hand, an antagonist of TRIM24 and BRPF1B, is the only compound at hand so far for the TRIM subfamily of BRD family VI (*85*).

As implied above, family VI contains two subfamilies, the RING-type E3 ubiquitin transferase of the TRIM family and bromodomain adjacent to zinc finger domain (BAZ2) family. Beside compound 34, two antagonists for the BAZ family have been developed, GSK2801 and BAZ2-ICR (*86, 87*). A characteristic of the BAZ family is a shorter ZA loop, which opens the binding pocket compared to other families (*59*).

The proteins of family VIII are part of SWI/SNF or Brg/Brahma-associated factors (BAFs) chromatin remodelling complexes. Mutations in these complexes are major factors in carcinogenesis and therefore compelling drug targets (*88*). PFI-3 is the only available potent inhibitor for this family with affinity towards the fifth of total 6 BRDs in Protein polybromo-1 (PB1) and the BRDs of SWI/SNF Related, Matrix Associated, Actin Dependent Regulator of Chromatin, Subfamily A, Member 2/4 (SMARCA 2/4). Interestingly, four conserved water molecules in the binding pocket are depleted upon inhibitor binding (*89*).

YEATS domains

Structure and extended recognition mechanism

Histone tail acetylation has been reported already in the 1960s (*90*) and numerous publications showed that these post-translational modifications play an important role in transcription control and the state of chromatin (*1*). Since then, other acylations have been discovered, comprising crotonylation, 2-hydroxyisobutyrylation, betahydroxybutyrylation, propionylation, butyrylation, succinylation, and glutarylation. Some of them compete with acetylation at the same site, others, such as crotonylation and 2-hydroxyisobutyrylation form specific modification patterns across the epigenome. They regulate spermatogenesis, early development, inflammation, and starvation. As mentioned above, short acyl chains are generated in the cytosol out of metabolic intermediates as acyl-CoA and therefore represent snapshots of the inner cell composition. HATs and HDACs regulate these modifications by accepting acyl-CoAs as substrates (*10*).

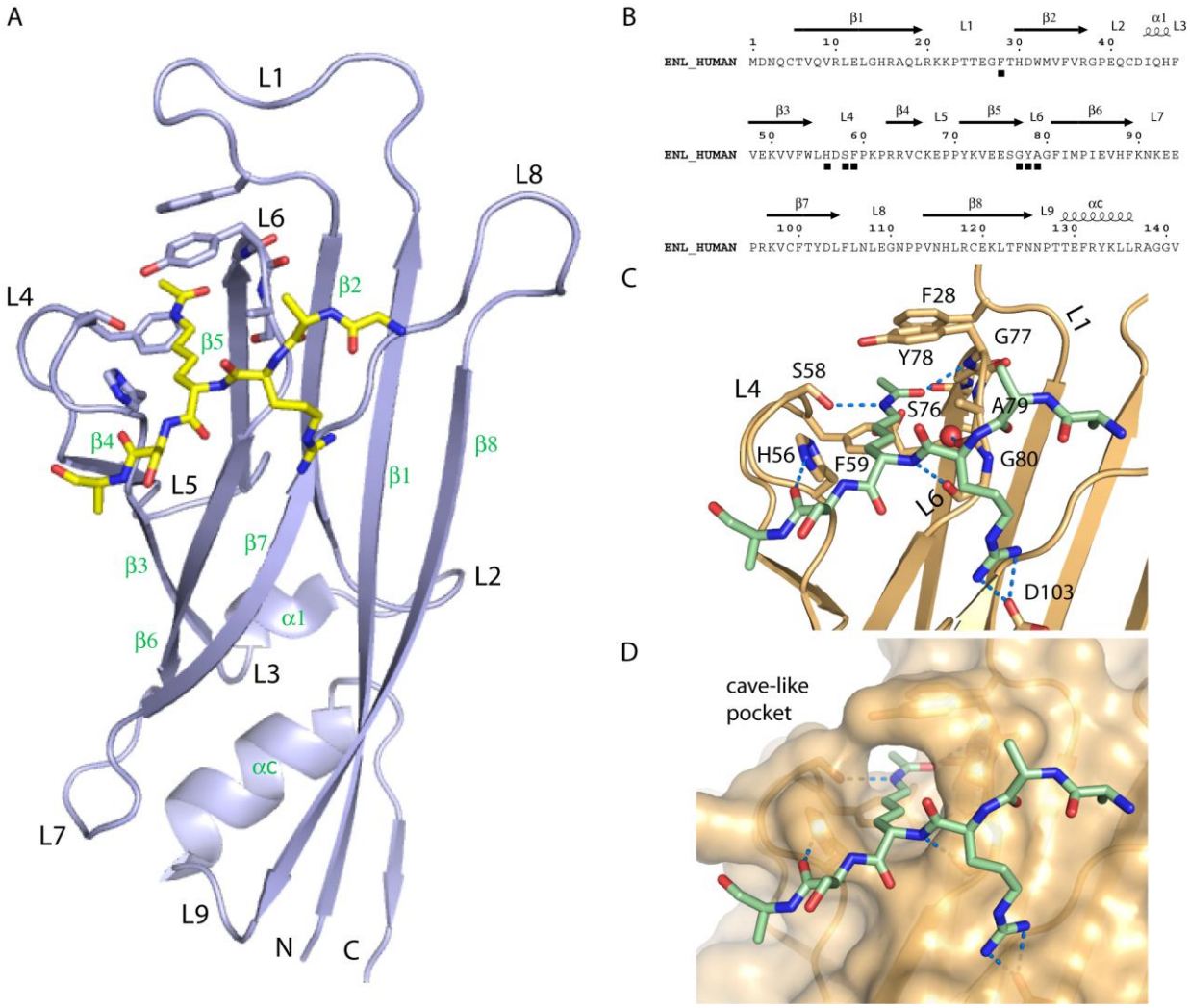


Figure 7 Overview over the YEATS domain of human ENL (PDB ID: 5J9S) (*19*). **A** YEATS domains adopt a beta-sheet fold with flexible loops emerging at both sides of the domain. Loops L4, L6 and L1 form the conserved binding pocket, which recognises histone peptides with acylated lysine sidechains. **B** Sequence of the ENL YEATS domain with annotated secondary structural elements. Partially illustrated with ESPript 3 (*61*) **C** Close-up of the ENL binding pocket with bound H3 peptide. Acetyl-lysine is recognised by an aromatic triad (F28-F59-Y78), the side chain of S58 in L4 and the amide backbone of L6. **D** Surface representation of the ENL YEATS domain. The binding pocket forms a cave-like structure that is open to the back.

YEATS domains have recently been identified as reader domains for acetylations and other acylations, linking the discovered histone acyl PTMs to other protein complexes (*10*). The structure of a YEATS domain with bound histone peptide is exemplified by ENL in complex with H3K27ac in Figure 7. The domain adopts an immunoglobulin fold with eight antiparallel beta strands and loops protruding at both heads and tails of the beta sheets (Figure 7A). The beta sheets and intermediate loops are numbered

chronologically from N to C terminus (Figure 7B). The binding pocket that recognises the modified histone peptide is formed by loops L4, L6 and L1. The acetyl-lysine moiety is sandwiched into an aromatic triad between F28 in L1, F59 in L4 and Y78 in L6 and interacts with the sidechain of S58 and the L6 amide backbone via hydrogen bonds. Three hydrogen bonds in close proximity to the binding pocket mediate the interaction towards the rest of the histone peptide. H56 in L4 and the G80 amide backbone of L6 form bonds to the histone peptide backbone and the D103 sidechain interacts with the histone arginine at position -1 relative to acetyl-lysine (Figure 7C). The surface representation emphasises the overall cave-like shape of the binding pocket, being elongated and open towards the back (Figure 7D) (*19*).

YEATS domains in health and disease

Four human proteins are known containing YEATS domains: ENL (also called myeloid/lymphoid or mixed-lineage leukemia translocated to chromosome 1 (MLLT1)), YEATS domain-containing protein 2 (YEATS2), ALL1-fused gene from chromosome 9 protein (AF9, also called MLLT3) and YEATS4 (also called Glioma-amplified sequence 41 (GAS41)) (Figure 8). AF9 was the first human YEATS domain shown to be a reader of both histone acetylations (*7*) and acylations with a preference for H3K9 crotonyl-lysine (*17*). Subsequently, histone acetylation reader activity was also confirmed for the AF9 paralogue ENL (*19*). Simultaneously, histone acylations were identified as preferential ligands for the YEATS2 binding pocket with the highest affinity towards H3K27 crotonyl-lysine (*18*). Although involvement of GAS41 in the NuA4 HAT complex linked the protein to histone acetylation some time ago (*91*), it was not until recently that evidence emerged for the GAS41 YEATS domain to function as pH-dependent H3K122 succinylation and crotonylation reader (*20*) as well as H3K27 and H3K14 acetylation reader (*92*).

AF9 and ENL are two similar proteins and are frequent fusion partners of MLL which give rise to a chimeric protein leading to acute myeloid leukemia (AML) with poor outcome. Both regulate transcription elongation in the super elongation complex, where they link histone acylation sites with transcription machineries by interacting with the scaffolding protein AF4/FMR2 family member 4 (AFF4) and RNA polymerase II-associated factor 1 homolog (PAF1) (*21, 93, 10, 94*). AF9 binds to H3K9 acetylation sites and promotes the methylation of H3K79 by interacting with DOT1L (*7*). The ENL YEATS domain was shown to be essential for progression of MLL-rearranged leukemia (*19, 21*). ENL also resides in the ENL-associated BAF-containing BAF250b (EBAF) complex of the SWI/SNF chromatin remodelling

complex families (*95*) (Figure 8). Furthermore, data supports that mutations in the ENL YEATS domain lead to the development of Wilms tumour in early renal development (*96*).

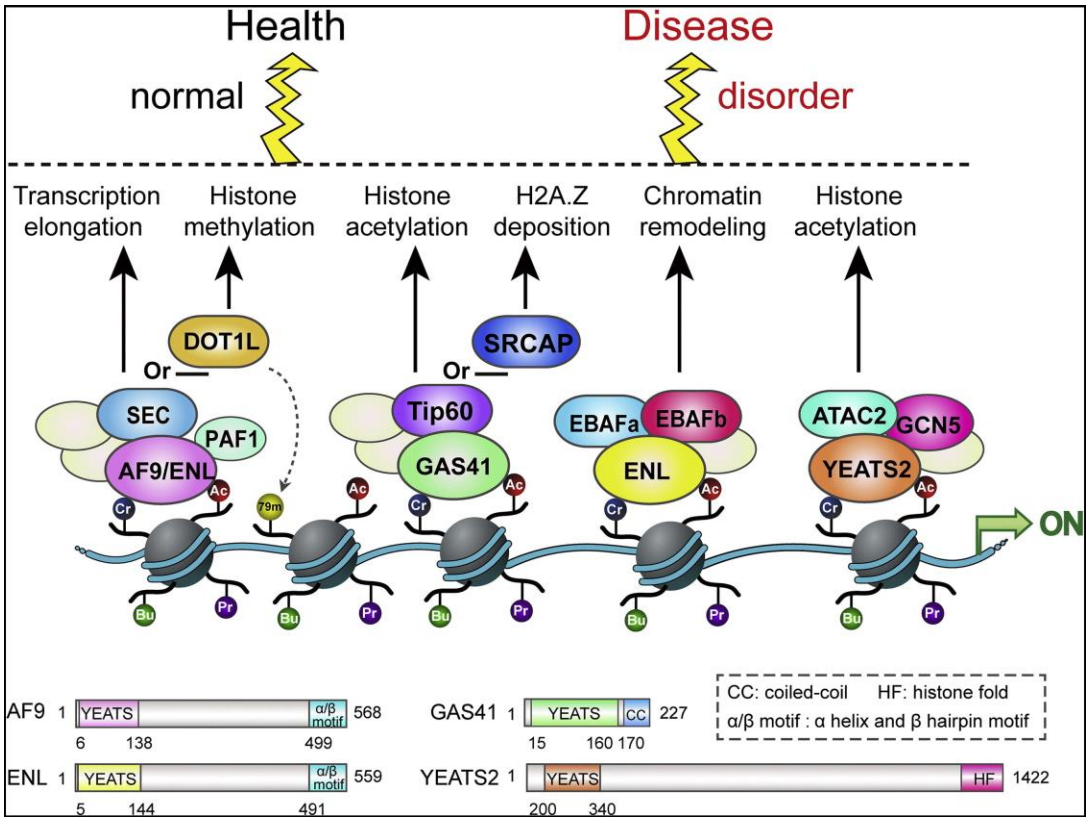


Figure 8 The four human YEATS domains and their function. AF9/ENL mediate the interaction between the super elongation complex (SEC) and Pol II through interaction with scaffolding proteins PAF1 and AFF4. YEATS2 is part of the GCN5 and ATAC HAT complexes. GAS41 regulates histone acylation and H2A.Z by either interacting with Tip60 or SRCAP complexes (adapted from (*10*), Copyright (2017), with permission from Elsevier)

YEATS2 is part of the GCN5 and human ADA2A-containing complex (ATAC) HATs. The YEATS2 gene is amplified in non-small cell lung cancer and is necessary for cancer cell growth and survival (*97*). In this context, YEATS2-mediated co-localisation of the ATAC complex to H3K9ac sites was investigated. The disruption of the YEATS-histone interaction also lead to the deactivation of essential genes (*10, 98*).

GAS41 is a subunit of the Snf2-related CBP activator protein remodelling complex (SRCAP) and the Tip60 histone acetyltransferase complex and regulates histone acetylation as well as H2A.Z deposition (Figure 8) (*38, 99*). Also, GAS41 is an oncogene and frequently amplified in various types of glioma, but it has also been linked to colorectal cancer cell proliferation and pancreatic cancer (*10, 100–102*). Finally, GAS41 interacts with the general transcription factor TFIIF (*103*).

Objectives

The ensemble of epigenetic modifications on histones, referred to as the histone code, regulates chromatin structure and gene transcription activity. BRDs and YEATS domains are epigenetic reader domains recognising histone tail acetylation and acylation marks. Several studies identified BRDs and YEATS domains as key players driving diseases connected to cancer, inflammation and metabolism, making them compelling drug target candidates (*12, 10, 104*).

In the recent years, the development of chemical probes for BRDs greatly contributed to the understanding of their physiological roles and their potential as pharmaceutical targets. So far, probes for seven out of eight BRD families are available, although some were predicted to have less druggability properties. Especially inhibition of family II, the BET family, results in pronounced physiological effects and influences a number of disorders and diseases. Their excellent druggability prompted not only the development of chemical probes, but also drugs that are now in clinical trials (*26, 25*). Some known kinase drugs were found to be dual BRD-kinase antagonists, on the hand explaining their long-lasting effects and on the other hand establishing a basis for rational dual inhibitor development (*14*). Despite these advancements, there are still many points unknown and worth investigating:

1. So far, 50% of all human BRDs are targeted by chemical probes, but most of them are designed for families II and IV. The examples of probes for ATAD2, BAZ2B, SMARCA2 and others showed that inhibitors can be engineered despite low druggability ratings (*12, 26*). Therefore, one aim of this thesis was to reveal additional characteristics in untargeted BRDs that enable the development of small molecule inhibitors. Also, despite the availability of many chemical probes, studies linking BRDs outside the BET family to phenotypes remained scarce. Thus, the elucidation of more BRD phenotypes was another goal of this thesis.

   In an effort to develop inhibitors for untargeted BRDs, a combination of ligand-based and structure-based virtual screening and docking techniques were used and optimised with the help of biophysical characterisation. Fragments were used in this approach to find new and modifiable chemotypes. The computational methods were deployed by collaboration partners Vassilios Myrianthopoulos and Emmanuel Mikros from the University of Athens, while biophysical measurements were contributed by me as part of the Stefan Knapp lab in Frankfurt am Main.

The discovery of dual BRD-kinase inhibitors explained why some drugs display less susceptibility to resistance and provided the basis for rational design of the same (*14*). The characterisation of rationally designed dual BRD-kinase inhibitors synthesised by collaborators at the Institute of Cancer Research (ICR, London) was another objective.

2. Neuroblastoma is a paediatric cancer and the most common extracranial solid tumour in childhood (*105*). Patients harbouring this disease share a mutated form of ALK, ALK^{F1174L}, and elevated MYCN levels (*106*). Unfortunately, published ALK inhibitors ceritinib and crizotinib do not inhibit the F1174L mutant efficiently. Since BRD4 is a regulator of the MYCN gene, compounds identified by Ciceri *et. al.* were used as basis to develop a new BRD4-ALK^{F1174L} dual inhibitor aimed at this type of aggressive cancer (*14*). Our collaborators Ellen Watts and Swen Hoelder from the Institute of Cancer Research (ICR) in London were responsible for the synthesis and *in vitro* kinase validation. Monika Raab and Klaus Strebhardt from the university medical centre Frankfurt am Main contributed their expertise in cellular phenotypical assays. My responsibility in this project were structural biology analyses of interactions between BRD4/ALK and the synthesised inhibitors by X-ray crystallography, *in vitro* characterisation on BRD4 by thermal shift assay and ITC and cellular target engagement assays towards ALK and BRD4 by NanoBRET.

YEATS domain lately emerged as versatile readers of histone acylations. All four human representatives were shown to be involved in diseases and in all cases, a direct relation between disease progression and reader domain activity was demonstrated (*38, 10*). Despite the promising reports about the relevance of YEATS domains in disease, the results were not yet verified by other publications and especially not with the use of chemical probes, when the research for this thesis started (March 2016).

3. Using a fragment-based and structure-based approach to target this new family of epigenetic reader domains, the general druggability of YEATS domains was to be validated. Furthermore, chemical scaffolds for inhibitors were to be developed and if feasible, optimised to be selective chemical probes. Encouraged by two simultaneous publications indicating a strong dependence of the ENL YEATS domain on the development of AML, this project focused on ENL and its relative AF9 (*19, 21*). While I utilised all biophysical means in our laboratory, our collaborators Moses Moustakim, Paul Brennan and Oleg Fedorov (University of Oxford) contributed their high-throughput and chemical synthesis efforts to structural and biochemical studies carried out in Frankfurt.

Materials and methods

Material lists

Table 1 Equipment

Equipment	Manufacturer
6230 TOF LC/MS; 1260 Infinity	Agilent Technologies
Analog Vortex Mixer	VWR
AREX DIGITAL PRO Heating Magnetic Stirrer	VELP®SCIENTIFICA
Benchtop UV Transilluminator	UVP
CLARIOstar plate reader	BMG Labtech
Digital Heatblock	VWR
DX-200 Autoclave	Systec
HERAFREEZE HFU T SERIES	Thermo Scientific
HighLoadTM 16/60 SuperdexTM 75/200	GE/Amersham Biosciences
HighLoadTM 16/600 SuperdexTM 75 pg	GE/Amersham Biosciences
Incubators (at 4 and 20 °C)	BINDER
Incu-Shake MIDI	SciQuip
LEICA M165 C; KL2500 LED; MC170 HD; PLANAPO 1.0x Microscope	LEICA
MEGA STAR 1.6R, 3.0R Centrifuge	VWR
MICRO STAR 17R Centrifuge	VWR
Mini-PROTEAN® Tetra System	BIO-RAD
Mosquito Pipetting robot	ttp labtech
Multichannel pipettes (0.5-10µL; 5-50µL; 30-200µL)	Starlab
Multitron Standard	INFORS HT
Nano ITC + 250 µL Titration syringe	TA Instruments
NANODROP 2000	Thermo Scientific
OLYMPUS SYX12; DF PLAPO 1x PF	OLYMPUS
pHenomenal® pH 1100 L	VWR
PhotoDoc-ItTM Imaging System	UVP
Power Source 250V	VWR
PRISMTM MicroCentrifuge	Labnet
Research plus Pipettes (1 ml, 200 µl, 20 µl, 1 µl)	Eppendorf
SORVALL LYNX 6000; T29-8x50 rotor; FiberliteTM F9-6x1000 LEX	Thermo Fisher Scientific

Table 2 Consumables

Consumable	Supplier
Microplate, 96 well, PS, V-Bottom clear, sterile	Greiner bio-one
96 well plate, U-bottom, 300 µl, Polypropylene	4titude
Thin wall 200 µl low profile PCR plate white	Starlab
3 lens low profile 96 well plate	SWISSCI
Pipet tips (10µL; 200µL; 1000µL)	Starlab
Serological Pipets (5mL; 10mL; 25mL; 50mL)	VWR

PCR Foil Seal, qPCR Seal	4titude
Amicon® Ultra (10kDa Cutoff) Centrifugal Filters (4mL, 15mL)	Merck Millipore Ltd.
StarTub Reagent Reservoir, PVC	Starlab
SnakeSkin® Dialysis Tubing 3,500 MWCO	Thermo Scientific
Pall filter Supor®-200 0.2µm 47mm	PALL Corporation
Thin Well 96 x 0.2mL Low Profile PCR Plates	Starlab
Reaction tubes (PP, graduated) (1.5mL, 2mL)	Greiner bio-one
Rotilabo®-Spritzenfilter PVDF, unsterile (0.22 µm; ø 13 mm)	Carl Roth

Table 3 X-ray crystallography equipment

Equipment	**Manufacturer**
Angled Cryo-Tong	Molecular Dimensions
Dry Shipper (CX100)	Molecular Dimensions
Equipment	Manufacturer
LD4 cryogenic dewar (4L capacity)	Molecular Dimensions
Magnetic CryoVials	Molecular Dimensions
Mounted CryoLoops	Hampton Research
Plain CryoCaps	Molecular Dimensions
Puck Dewar Loading Tools	Molecular Dimensions
Puck Loading Dewar	Molecular Dimensions
Puck Separator Tools	Molecular Dimensions
Puck Wand	Molecular Dimensions
Puck-Shelved Shipping Cane	Molecular Dimensions
Shipping case for CX100 Dry Shipper	Molecular Dimensions
SPINE Pucks	MiTeGen
Uni-Pucks	Molecular Dimensions

Table 4 Chemicals and ready-to-use solutions

Material	**Manufacturer**
Acetonitril	VWR chemicals
Acrylamide/bis 40%	G Bioscienes
Agar	amresco
Ammonium persulfate	Sigma-Aldrich
Ampicillin	Carl Roth
Chloramphenicol	amresco
Coarse screens (compositions in Supplementary Tables S2-S6)	Molecular Dimensions
Dimethyl sulfoxide	Carl Roth
D-Sucrose	Fisher
Ethylene glycol	Sigma-Aldrich
Formic acid	VWR chemicals
Glycerol	Carl Roth
Glycerol	Carl Roth
HEPES	Fisher

Imidazole	Alfa Aesar
IPTG	VWR chemicals
Isopropanol	Carl Roth
K_2HPO_4	amresco
Kanamycin	Carl Roth
KH_2PO_4	VWR Chemicals
LB medium	Carl Roth
Nickel(II) sulfate hexahydrate	Sigma-Aldrich
Nε-acetyl-L-lysine	Sigma-Aldrich
Polyethyleneimine	Alfa Aesar
Precision Plus Protein Unstained Marker	BIO-RAD
Quick Coomassie Stain	generon
Sodium chloride	Amresco
Sodium hydroxide	Carl Roth
SYPRO orange dye	Invitrogen
TCEP	Zellbio
TEMED	Carl Roth
Tris base	Fisher
Tryptone	Carl Roth
Yeast Extract	MELFORD Biolaboratories Ltd.

Table 5 Buffers and media

Buffer	**Composition**	**Amount**
4x TB	Yeast extract	24 g/l
	Tryptone	12 g/l
	Glycerol	4 ml/l
	KH_2PO_4	0.017 M
	K_2HPO_4	0.072 M
LB medium	LB powder	25 g/l
LB agar	LB powder	25 g/l
	Agar	7.5 g/l
BRD lysis buffer	HEPES pH 7.5	25 mM
	NaCl	500 mM
	TCEP	1 mM
	Glycerol	5% (v/v)
	Imidazole	5 mM
BRD elution buffer	HEPES pH 7.5	25 mM
	NaCl	500 mM
	TCEP	1 mM
	Glycerol	5% (v/v)
	Imidazole	300 mM
BRD gel filtration buffer	HEPES pH 7.5	25 mM
	NaCl	300 mM
	TCEP	0.5 mM
	Glycerol	5% (v/)
BRD DSF buffer	HEPES pH 7.5	10 mM
	NaCl	500 mM
BRD ITC buffer	HEPES pH 7.5	10 mM

	NaCl	150 mM
	TCEP	0.5 mM
	Glycerol	5% (v/v)
ENL lysis buffer	Tris pH 7.5	25 mM
	NaCl	500 mM
	TCEP	1 mM
	Glycerol	5% (v/v)
	Imidazole	5 mM
ENL elution buffer	Tris pH 7.5	25 mM
	NaCl	500 mM
	TCEP	1 mM
	Glycerol	5% (v/v)
	Imidazole	300 mM
ENL gel filtration buffer	Tris pH 7.5	25 mM
	NaCl	300 mM
	TCEP	0.2 mM
ENL DSF buffer	HEPES pH 7.5	10 mM
	NaCl	150 mM
ENL ITC buffer	Tris pH 7.5	25 mM
	NaCl	500 mM
	TCEP	0.5 mM
	Glycerol	5% (v/v)
SDS-PAGE running buffer 10x	Tris base	0.25 mM
	Glycine	1.92 M
	SDS	1% (w/v)
SDS-PAGE loading dye 4x (20 ml)	Glycerol	8 ml
	Tris pH 6.8 1M	4 ml
	DTT 1 M	8 ml
	SDS	0.6 g
	Bromphenol blue	Spatula tip

Table 6 Software

Software	**Provider**
BIOVIA Draw BIOVIA	BIOVIA Draw BIOVIA
Office Suite 2016	Microsoft
GraphPad Prism 5	GraphPad Software, Inc.
ICM-BrowserPro v3.8.6	MolSoft L.L.C.
Leica Application Suite v4.8.0	Leica Microsystems
MxPro (Mx3005P) v4.10	Stratagene
NanoAnalyze TA	NanoAnalyze TA
NanoDrop 2000 (Protein A280) Thermo Scientific	NanoDrop 2000 (Protein A280) Thermo Scientific
PrimeView 5.31; PrimeView 5.31 Evaluation	PrimeView 5.31; PrimeView 5.31 Evaluation
General Electric Company	General Electric Company
MassHunter BioConfirm	Agilent
CCP4 Suite 7.0	(*107*)
COOT 0.8.9	(*108*)
PyMol 1.7.6.2	Schrödinger L.L.C.

Molecular cloning

Ligation-independent cloning (LIC)

Gaining structural information on protein targets of medical relevance is a main goal of the Structural Genomics Consortium (SGC). The crystallisation of new proteins requires generating a great number of genetic constructs which include varying parts of the protein sequence. Despite of useful prediction methods for globular parts and disordered regions (*109, 110*), it is necessary to test different sequences as well as different affinity tags at either N- or C-terminus of the sequence.

LIC is a cloning method that works without the use of T4 DNA ligase, making it suitable to generate libraries of clones in a high-throughput manner (*111*). Its principle is based on overlapping and complementary 5'-single-stranded (ss) tails on both the vector and the insert. The primers used in the polymerase chain reaction (PCR) (*112*) to create the insert contain an additional 15-nucleotide 5'-sequence lacking deoxycytidine monophosphate (dCMP). During the PCR, a complementary 3'-sequence with lacking deoxyguanosine monophosphate (dGMP) is added, which is removed by the 3'-5' exonuclease activity of the T4 DNA polymerase in the presence of deoxyguanosine triphosphate (dGTP) to yield 5'-extending single-stranded (ss) tails. The vector plasmids, in this case pNIC28-Bsa4 and pNIC-CH (Figure 9), contain 15-nucleotide sequences complementary to those on the insert, but with lacking dGMP on the 5' extension. Similarly, ss-tails on the vector are produced by using the same T4 DNA polymerase activity in the presence of deoxycytidine triphosphate (dCTP). The preparation of the vector and inserts with complementary 5' ss-overlaps prevents on the one hand that the vector closes without insert and on the other hand allows to use the vector-plasmid mixture directly for bacterial transformation. Selection of bacteria with plasmids containing inserts is performed by use of sucrose in connection with the *sacB* gene from *B. subtilis*. Due to this gene, bacteria express levansucrase, leading to lethal amounts of levan in the periplasm, unless the gene is perturbed by the insert (*113, 114*).

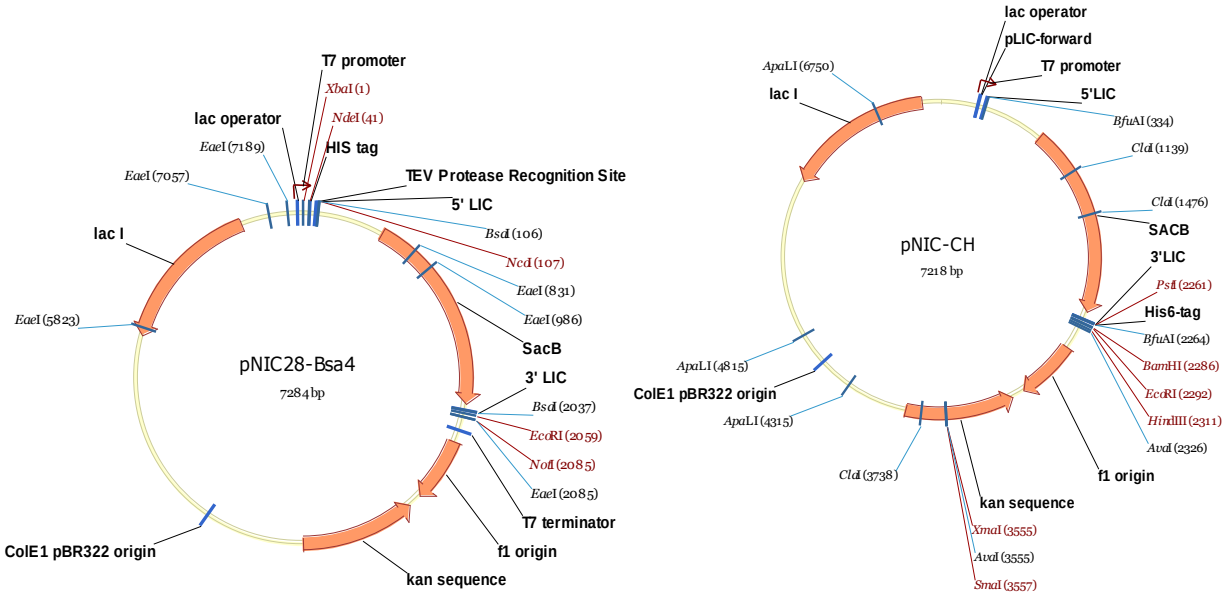


Figure 9 Vector maps for pNIC28-Bsa4 and pNIC-CH. Both vectors contain dGMP-free 15-nucleotide sequences suitable for LIC and allow *lac* operon-controlled gene transcription with the T7 polymerase. pNIC28-Bsa4 contains a cleavable hexahistidine tag (His_6 tag) sequence N-terminal to the protein sequence. pNIC-CH contains a non-cleavable His_6 tag C-terminal to the protein sequence.

Protocol PCR and cloning

First, forward and reverse primers are mixed to a 5 µM concentration. Template DNA is diluted to approx. 2.5 ng/µl. 21 µl of master mix (Table 7) are added to PCR tubes, followed by 1.5 µl primer mix and 2.5 µl template DNA. The tubes are put into a thermocycler and a program with following conditions is run (tempX is the average melting temperature of the primers minus 10 °C):

94° C, 5 min
(94° C, 30 sec; tempX, 30 sec; 68° C, 1 min) x25 cycles
72° C, 10 min
15° C hold

Table 7 Master mix (x10) for PCR reaction

Component	Amount
HercII buffer (5x stock)	50 µl
10 mM deoxynucleoside triphosphate (dNTP) mix	7.5 µl
HercII polymerase	2.5 µl
Water	150 µl

The resulting fragments are analysed by gel electrophoresis. In the meantime, the vector is digested in a total volume of 100 µl by mixing 5 µg plasmid with 10 µl CutSmart buffer and 2 µl of BsaI restriction enzyme and incubating for 2 h/37° C. The insert after PCR is treated with DpnI enzyme, which removes the original template DNA from the solution and is then purified with a PCR purification kit. T4 DNA polymerase treatment of vector and insert is run for 30 min/22° C and the reaction is stopped for 20 min/75° C (Table 8).

Table 8 Reaction mixtures for T4 DNA polymerase treatment of vector and insert

Vector	**Insert**
21.5 µl water	5 µl DNA
50 µl BsaI-digested plasmid	2.15 µl water
10 µl NEB Buffer 2.1	1 µl 10x NEB2.1 buffer
10 µl dGTP (25 mM)	1 µl dCTP (25 mM)
1 µl BSA (NEB)	0.1 µl BSA (NEB)
5 µl DTT (100 mM)	0.5 µl DTT (100 mM)
2.5 µl NEB T4 polymerase	0.25 µl NEB T4 polymerase

1 µl of treated plasmid with 2 µl of treated insert are mixed, incubated for 30 min at RT and used for transformation. Insert-containing clones are selected on LB-Agar plates with 50 µg/ml kanamycin and 5% sucrose. Furthermore, insertion is validated by colony PCR using MyTaq DNA polymerase (Table 9) and the following thermocycler settings:

94° C, 10 min
(94° C, 30 sec; 50° C, 30 sec; 72° C, 30 sec) x25 cycles
72° C, 5 min
15° C hold

Table 9 Master mix (x10) for colony PCR reaction

Component	**Amount**
Bioline buffer (5x stock)	40 µl
pLIC-for/pLIC-rev primer mix (10 µM each)	10 µl
Bioline MyTaq DNA polymerase	1 µl
Water	149 µl

Protein expression and purification

Transformation

In order to transform plasmid DNA into *E. coli* cells, chemical competence and a heat shock are utilised. Chemical competence is achieved by neutralising negative repulsive forces between the negatively charged cell membrane and the DNA phosphate backbone by incubating the cells with mono-/bivalent anions. Membrane permeability and DNA uptake is increased further by a heat shock and subsequent incubation on ice, which presumably facilitates the formation and closure of pores (*115, 116*).

Protocol Transformation

50-100 ng DNA are added to 50 µl competent cells and incubated for 15 min on ice. In case of plasmid amplification, Mach1 cells are used. In case of protein expression, BL21 (*DE3*) R3 pRARE2 cells are used, which are resistant to phage contamination (R3) and contain an additional plasmid coding for transfer (tRNA) with optimised codon usage for human genes (pRARE2) and chloramphenicol resistance. Next, the cells are heat shocked at 42°C/45 s in a water bath and incubated for 3 min on ice. 100 µl of fresh 2xLB medium without antibiotics are added to the cell suspension and the cells are incubated at 37°C/1h. 30 µl of resuspended cells are stroke out onto LB-Agar plates with 50 µg/ml kanamycin and 34 µg/ml chloramphenicol for selection. The colonies are grown at 37°C/ON in an incubator.

Protein expression in BL21 system

Recombinant protein is overexpressed in the *E. coli* BL21 (*DE3*) expression system. In this system, the bacteriophage T7 RNA polymerase is encoded as *DE3* gene in the bacterial genome (*117, 118*), which is controlled by a lacUV5 promoter. Upon transcriptional activation by Isopropyl-β-D-thiogalacto-pyranosid (IPTG), a non-metabolisable analogue of lactose, T7 RNA polymerase is expressed and mediates the transcription of the gene of interest by specific binding to the T7 promoter on the previously transformed plasmid vector. The advantage of the T7 system is its tight promoter specificity and high transcription rate (*119*). Induction by IPTG is conducted in the exponential growth phase of the culture, favouring cell environments that facilitate protein expression. After induction, cell growth and protein integrity are controlled by temperature, shaking frequency and overall medium volume.

Protocol Protein expression

A pre-culture consisting of 2xLB medium with 50 µg/ml kanamycin and 34 µg/ml chloramphenicol is inoculated with a single colony from an LB-Agar selection plate and grown overnight at 37° C/250 rpm. Next day, 10 ml of pre-culture are used per 1 l TB medium large-scale expression culture for inoculation

and the culture is shook at 37° C/180 rpm. When the optical density (OD) at 600 nm reaches 1.5, temperature is reduced to 18° C. At on OD of 3.0, protein expression is induced with 0.5 mM IPTG and the culture is grown at 18° C/180 rpm overnight. The bacterial cells are harvested at 4° C/8,700xg/15 min and stored at -20° C before proceeding.

Immobilised metal affinity chromatography (IMAC)

Affinity tags are used to clear the protein of interest from cell lysate as first step of protein purifications. A frequently used setup is an N-terminal or C-terminal His_6-tagged protein and a gravity column loaded with nickel-nitrilotriacetic acid (Ni^{2+}-NTA). NTA chelates Ni^{2+} ions and is linked to agarose beads, which form the column bed. During purification, close proximity histidine side chains (as in the affinity tag) bind to a Ni^{2+} ions, while the rest of the solution elutes from the column. The separated protein of interest is retrieved from the column by applying imidazole-containing buffer, which competes with the histidine side chains in a concentration-dependent manner. Alternatively, Co^{2+}-NTA agarose can be used instead of Ni^{2+}-NTA, which theoretically increases purity but decreases yield, because Co^{2+} does not chelate histidine side chains as effectively.

Protocol for bromodomain IMAC

Cells containing His_6-tagged protein are suspended in BRD lysis buffer (buffers in Table 5, page 41) and lysed (3 min pulse time) on ice using sonication. After that, polyethylene imine at pH 7.0 is added to precipitate cellular DNA. The lysate is cleared by centrifugation at 23,000 rpm/45 min/4° C and applied to 5 ml Co^{2+}-NTA agarose column equilibrated with 10 column volumes (CV) BRD lysis buffer. After reapplying to the column, the resin is washed with 10 CV of BRD lysis buffer at gravity flow and the protein is eluted using a step elution of imidazole in 2 CV of BRD lysis buffer (20 mM, 50 mM, 100 mM, 2x300 mM). All fractions are collected and monitored by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS‑PAGE).

The His_6 tag was cleaved in for crystallisation experiments with bromodomains as described in the following paragraph:

The eluted protein is filled into a dialysis tube with 3.5 kDa molecular weight cutoff (MWCO). Tobacco etch virus (TEV) protease is added at 1:20 molar ratio (protein:TEV) and the solution is incubated at 4° C overnight while mixing on a magnetic stirrer at 100 rpm. Next day, the solution is depleted from TEV protease and the cleaved His_6 tag through rebinding to 2 ml Ni^{2+}-NTA agarose equilibrated with 5 CV of BRD gel filtration buffer. Remaining protein is eluted in two steps using BRD gel filtration buffer with

30 mM imidazole and BRD gel filtration buffer with 300 mM imidazole. Samples are monitored by SDS-PAGE.

Protocol for ENL YEATS IMAC

Cells containing His_6-tagged protein are suspended in ENL lysis buffer and lysed (3 min pulse time) on ice using sonication. After that, polyethyleneimine at pH 7.0 is added to precipitate cellular DNA. The lysate is cleared by centrifugation at 23,000 rpm/45 min/4° C and applied to 2 ml Ni^{2+}-NTA agarose column equilibrated with 10 column volumes (CV) ENL lysis buffer. After reapplying to the column, the resin is washed with 25 CV of ENL lysis buffer, then 10 CV of ENL lysis buffer with 30 mM imidazole at gravity flow and the protein is eluted using 2x4 CV of ENL elution buffer. All fractions are collected and monitored by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE).

Fractions containing protein of interest are subjected to size exclusion chromatography (SEC) as final purification step.

Size-exclusion chromatography

SEC is a method to analyse and purify macromolecules. It is based on separation over a column with porous beads depending on molecular weight, or more specifically, the hydrodynamic radius of molecules. First attempts were made with conventional starch (*120*), whereas nowadays crosslinked dextran polymers, agarose or acrylamide are used. Smaller molecules move through more pores in the beads, resulting in slower retention time. Larger molecules do not enter the pores as much and therefore move faster through the column (*121*). Analytes are pumped through the column and do not chemically interact with the beads. The relative retention factor for a substance follows the equation

$$(1)\ \mathrm{Kav} = \frac{V_e - V_0}{V_t - V_0},$$

whereas V_e represents the specific retention volume, V_t the total column volume and V_0 the void volume.

Protocol Size-exclusion chromatography

Prior to sample preparation, the column and the sample loading loop are equilibrated with gel filtration buffer. The elution fractions after IMAC or after rebinding to Ni^{2+}-NTA are concentrated to 5 ml, centrifuged at 17,000xg/5 min/4° C and loaded into the loop through a 0.22 µM filter. Next, the pump system is washed with gel filtration buffer and the sample volume is injected onto the column. Protein-containing fractions are confirmed by UV absorption at 280 nm, SDS-PAGE and mass spectrometry.

Samples with suitable purity are pooled and concentrated in 10 kDa MWCO centrifugal filters before shock-freezing in liquid nitrogen or direct use for crystallisation experiments. For a list of gel filtration buffers, refer to table 5, page 41.

Protein analytics

Determination of protein concentration

The concentration of a protein sample is determined by measuring its absorption at 280 nm. The specific molar absorption of an unfolded protein is proportional to its relative tyrosine and tryptophan content (and to lesser extent also phenylalanine and disulphide bonds) and can be calculated judging the protein sequence (*122*). The resulting extinction coefficient ε_λ is part of the Beer-Lambert law, which correlates absorption with concentration:

$$(2)\ E_\lambda = \varepsilon_\lambda * c * d\ ,$$

where E_λ describes the extinction, c the concentration and d the path length of the beam through the sample (*123*). For proteins with no or very few aromatic side chains, the concentration can be calculated alternatively via absorption of the protein backbone at 205 nm (*124*).

Protocol Determination of protein concentration

The theoretical extinction coefficient ε_λ and the protein mass are input into the software. A background spectrum is measured with 1.5 µl of the respective protein buffer. The device is cleaned, and a protein spectrum is measured with equal volume. The protein concentration is calculated by the software according to equation 2.

SDS-PAGE

SDS-PAGE is a method for separating proteins depending on a size by applying an electric current. The key to this method is the use of SDS, a detergent, which denatures the protein and applies a negative charge to the molecule that is proportional to its molecular weight. Polyacrylamide gels are composed of a stacking gel and a separating gel, while the latter can be prepared at different concentrations. The higher the concentration, the better the separation resolution at lower molecular weights and *vice versa*. Other components of the loading dye include bromophenol blue, making the sample visible in

buffer, glycerol, keeping the sample in the loading pocket of the gel, and β-mercaptoethanol, reducing disulphide bonds.

After preparing the sample by mixing it with loading dye and heating it, the sample is applied to a polyacrylamide gel, more specifically to the stacking gel half with pH 6.8. The purpose of the stacking gel is to compress all protein components in a thin band before it enters the separating gel. This is achieved with glycine and chlorine ions in the buffer. Glycine, being almost neutral at this pH, forms an electric gradient with the chlorine ions, which cumulates all protein molecules into a thin band. After entering the separation gel which has a pH of 8.8, glycine will carry a negative charge, moves ahead in the gel with the chloride ions and leaves all protein molecules to migrate according to their molecular weight (*125*). Subsequent staining of the gel and comparison with a marker containing proteins of known size allows different protein components in the samples to be judged in size and purity.

Protocol SDS-PAGE

Resolving and stacking gel components (Table 10) are mixed separately in Erlenmeyer flasks without ammonium persulfate (APS) and tetramethylethylenediamine (TEMED). Polymerisation of the resolving gel is induced by adding APS and TEMED quickly and the liquid is pipetted into a Multi-Casting chamber. The liquid is covered with isopropanol to avoid oxidisation. After polymerisation, the isopropanol layer is depleted, APS and TEMED are used to the stacking gel solution and the liquid is pipetted on top of the separation gel. Combs are inserted to form pockets. The gels are wrapped in wet tissue and stored at 4° C until use.

Protein samples are mixed with loading dye and incubated at 95° C for 1 min. Meanwhile, the gel is fixed into clamps and put into the running chamber. The running chamber is filled with running buffer and the heated protein samples are pipetted into the gel pockets. The chamber is connected to a power supply and electrophoresis is performed for 45 min at 200 V. Finally, the gel is shortly rinsed with water and stained with instant dye for 15 min or longer.

Table 10 Composition of SDS-PAGE gels

Component	**Separation gel 15% [ml]**	**Stacking gel**
Water	28.4	12.4
40% acrylamide/bis	30	4.3
1.5 M Tris pH 8.8	20	-
1.0 M Tris pH 6.8	-	4.2
10% SDS	0.8	0.34
10% APS	0.8	0.3
TEMED	0.032	0.03
50		

Electrospray ionization time of flight (ESI-TOF) mass spectrometry

ESI-TOF allows the determination of protein mass, being far more accurate than SDS-PAGE and therefore also detecting mutations, modifications like phosphorylation and others (*126*). It is especially useful as quality control for crystallisation experiments, as homogeneity of the sample is crucial. The setup contains two core elements: an ionization chamber for transforming the macromolecules into an ionized aerosol and the TOF unit for determination of flight times and respective mass/charge ratios.

The advantage of ESI is that it is a "soft" ionising method which avoids severe fragmentation of protein molecules. Proteins are injected into a spray chamber through a thin needle under slow flow (nl/min to µl/min). By applying high voltage at the needle, the injected is nebulised to form charged droplets. Further vaporisation takes place when injecting warm neutral gas. Two models describe the transition into the gas phase while the droplets are driven to the source (Figure 10). In the ion evaporation model, repulsive forces between the charged ions increase during the vaporisation, eventually exceeding surface tension to form single molecules without solvent (*127*). In the charge residue model, single charged molecules in the gas phase are formed by alternating solvent evaporation and fragmentation of droplets until no solvent molecules are present (*128*).

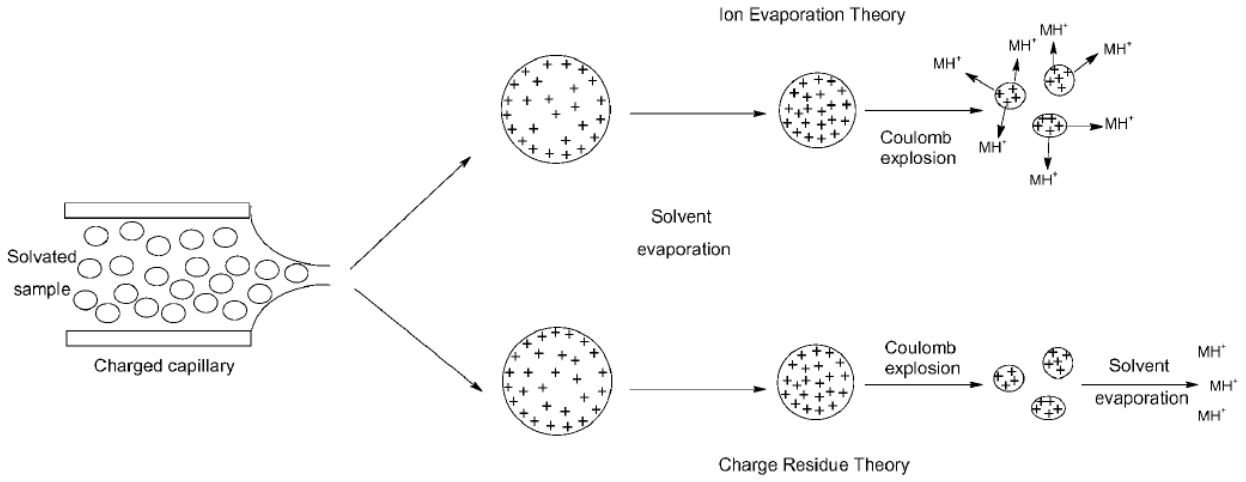


Figure 10 Formation of charged ions in ESI. In the ion evaporation theory, repulsive forces lead to transition into the gas phase. In the charge residue theory, solvent evaporation drives the transition (adapted from (*129*), Copyright (2009), with permission from Taylor & Francis).

TOF is one way to separate ions with different m/z (mass to charge) ratios. In the classical setup, it consists of a tube of 1-2 m length with the ions entering the tube on one side and a detector on the other side. Driven by voltage, the ions travel through the tube with different speed depending on their

mass and charge. Molecules with smaller mass and equal charge will reach the detector earlier, following equation

$$(3)\ m/z = \frac{t_f^2 2Es}{2s+x},$$

where t_f is the time of flight, E the applied voltage, s the acceleration length and x the length of the free flight.

A mass spectrum displays the m/z ratio on the x-axis and the detector count on the y-axis. Typically, protein molecules will exhibit different charge states per species in the spectrum, leading to a so-called charge envelope. The position of the charge envelope in the spectrum depends on the surface accessible area of the protein. Proteins with higher accessible area will get more charges per mass and therefore appear at lower m/z values. Since the peaks originate from one species with different integer charges, an envelope can be translated to an accurate mass by dividing the peaks with integer numbers, a process called deconvolution. The resulting graph displays intensity on the y-axis and mass on the x-axis. Per protein species, there will be one main peak with additional smaller peaks present, the latter arising from sodium adducts.

Protocol ESI-TOF

For ESI-TOF protein measurements, a setup containing Agilent 6230 ESI-TOF mass spectrometer and a 1260 Infinity LC unit with a C3 column is used. Protein samples are diluted to 0.1 mg/ml in 0.1% formic acid. 60 µl of the solution are pipetted into a 96 well U-bottom in the autosampler unit. Samples are measured in positive mode, with 10 µl injection volume and 0.4 ml/min flow rate. The chromatography gradient between 0.1% formic acid in water and 0.1% formic acid in acetonitrile is chosen according to Table 11. Before and after each sample run, a blank (0.1% formic acid) is measured to wash the chromatography column and detect possible contaminants. Agilent MassHunter BioConfirm is used to select peaks of the chromatogram and deconvolute the respective mass spectra.

Table 11 Gradient settings chromatography before ionisation

Time [min]	Water + 0.1% FA [%]	Acetonitrile + 0.1% FA [%]
0.00	97.0	3.0
1.20	95.0	5.0
3.00	15.0	85.0
4.00	5.0	95.0
7.50	5.0	95.0
9.50	95.0	5.0
10.50	95.0	5.0

Assays

Differential scanning fluorimetry (DSF)

DSF is a high-throughput screening (HTS) method to identify inhibitor hits for target proteins. It uses SYPRO orange dye, which is fluorescent if bound to hydrophobic moieties. It absorbs light at 492 nm and excites at 610 nm, a range which does not interfere with protein backbone or sidechain fluorescence. In the experiment, protein samples are gradually heated from room temperature to 96° C. During the denaturing of the protein samples, SYPRO orange binds to the protein, which leads to an increase of fluorescence. Reaching even higher temperatures, the protein aggregates and the dye dissociates, resulting in a decrease of fluorescence. By determining the point of change in the increasing fluorescence, one can identify the melting point T_M of a protein. This is achieved by fitting the increasing part of the graph with a Boltzmann equation (*130*):

$$(4)\ y = LL + \frac{UL-LL}{1+\exp\left(\frac{T_M-x}{a}\right)},$$

where LL and UL describe minimum and maximum intensities and *a* stands for the slope of the curve at T_M.

Compounds with affinity to the target will stabilise its structure and, in many cases, increase its melting point T_M. By subtracting the T_M of the control measurement target with DMSO from the T_M of the target with compound, one can calculate the so-called thermal shift. However, although thermal shifts usually correlate with affinities, absolute values are only comparable within measurement for one target. Target with higher flexibility, e.g. kinase domains, tend to give higher thermal shifts than the smaller bromodomains or thermally stable YEATS domains.

Protocol DSF

Protein samples are prepared at 2 µM final concentration in BRD DSF buffer or ENL DSF buffer. Then, SYPRO orange dye is added 1:1000, resulting in 5x final concentration. If proteins are thermally stable or if handling with low affinity compounds, the salt concentration in the DSF buffer is reduced to 150 mM. Compounds are prepared in 96 well plates with 500 µM concentration in DMSO.

In a 96 well low-profile PCR plate, 20 µl of protein solution with dye are aliquoted into each well. Next, 0.4 µl of compound from the source plate are added to the wells and mixed with the pipette. This will give a final compound concentration of 10 µM or 5x compound excess, respectively. The plate is sealed with qPCR seal and placed into a Stratagene Mx3005P qPCR system. In the operating software MxPro

v4.10, the SYBR green setting is chosen and the program is set to increase temperature from 25° C to 96°C with 3° C/s. After the measurement, data points are exported in excel files and the sigmoidal parts of the melting curves are imported into GraphPad for fitting with equation 4.

Isothermal titration calorimetry (ITC)

ITC is a method to determine thermodynamic parameters of binding events, typically between protein and ligand or protein and protein. The instrument detects heat changes which are evoked by enthalpic contributions of the binding event. The quantification of binding heat is achieved by comparison of two identical cells, a sample cell and a reference cell, both equipped with sensitive thermal units (Figure 11). Every instrument has a certain equilibrium heat rate that is historically measured in µcal/s. Through a syringe, a titrant is titrated into the sample cell in increments, which either leads to an exothermic or an endothermic reaction. The thermal units then detect a temperature difference between sample cell and reference, which is compensated by a feedback heating response with adjusted heating rate until the instrument reaches equilibrium. This leads to characteristic peaks in the ITC data if the heating rate is plotted against time.

In a typical measurement, binding between protein and an inhibitor leads to formation of e.g. hydrogen bonds, being detected as exothermic output that is compensated by cooling of the sample cell. Thus, negative peaks corresponding to “negative heating” are visible in the plot. In the first titrations of the experiment, all titrant molecules interact with their binding partners, leading to constant heat integrals per titration. Further titrations will gradually lead to smaller integrals because less and less titrant molecules bind to their partner molecules, leading to a sigmoidal slope and finally to saturation. The heat that is produced at saturation concentrations are evoked by sole dilution of titrant molecules in the sample cell.

Plotting the peak integrals against the molar ratios of the binding partners delivers information to calculate all thermodynamic parameters. The heat integral of the first titrations corresponds to ΔH, while the slope at the turning point of the sigmoidal curve refers to the dissociation constant K_D. The steeper the transition, the higher the affinity, because most molecules will bind per titration until saturation is reached. The Gibbs free energy difference ΔG and entropy difference ΔS are calculated by equations 5 and 6 and therefore interdependent with the measured enthalpy and dissociation constant.

$$\Delta G = RT \ln K_D \quad (5)$$

$$\Delta G = \Delta H - T\Delta S\,, \quad (6)$$

where R stands for the universal gas constant.

Protocol ITC

Measurements between protein and inhibitors are run in a so-called reverse titration with the protein in the syringe at 200-300 µM and the inhibitor in the sample cell at 10-50 µM on a TA Instruments Nano ITC device. The syringe is rotated at 350 rpm until the instrument reached the equilibrium heating rate. Titrations for bromodomains are run at 15° C and for ENL YEATS at 37° C. The first injection at 4 µl primes the sample cell, before 29 consecutive 8 µl injections at 200 s intervals comprise the main experiment. The peak integrals are plotted against the molar ratio between protein and inhibitor and fitted with a Boltzmann equation to determine K_D and ΔH. Respective buffers are listed in Table 5, page 41.

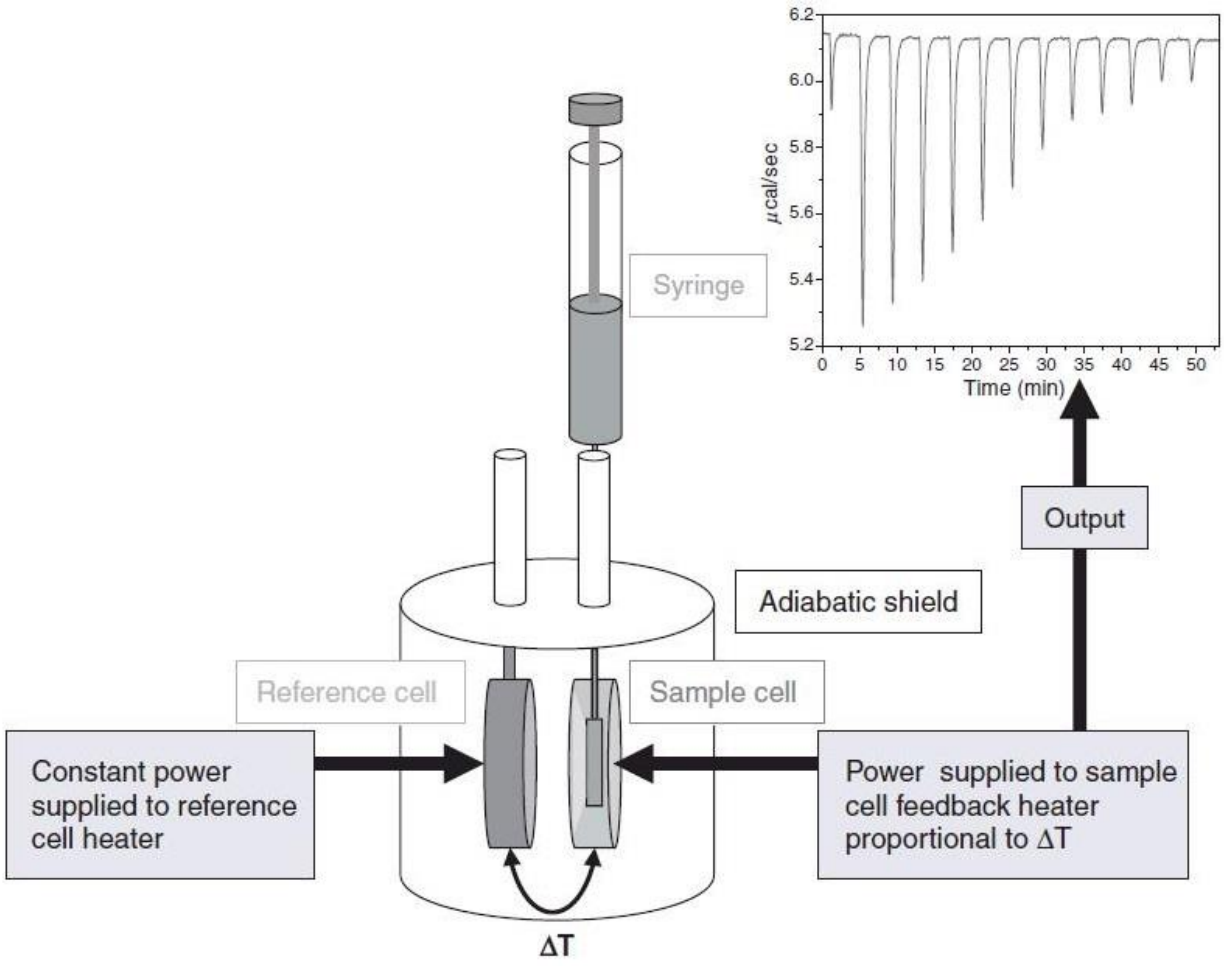


Figure 11 Setup of an isothermal titration calorimeter. One binding partner in the syringe is titrated to its counterpart in the sample cell, where their interaction changes the cell temperature compared to the reference cell. The compensation of this temperature difference yields the typical ITC raw heat plots which depict heating rate against time. (adapted from (*131*), Copyright (2008), with permission from Elsevier)

NanoBRET

A property in the inhibitor development that must be taken into consideration is its ability to bind the target protein in cells, also called target engagement. Measuring target engagement is relevant, because classical biophysical assays often do not provide information on several important influences on inhibitor activity, for instance the behaviour of the protein if expressed as full-length protein, its chemical environment in the cellular compartment as well as cell permeability and off-target effects of the inhibitor. Furthermore, by adding the cellular context, a more accurate view on the competition with the natural substrate/ligand of the target protein can be obtained. For example, the competition with adenosine triphosphate (ATP) in case of kinases has generally a great effect due to the high cellular concentration and different ATP affinities towards different kinases. A way to quantify target inhibition in cell in a dose-dependent manner is bioluminescence resonance energy transfer (BRET). In this thesis, an optimised setup with a bright luciferase (Nanoluc/Nluc) is used, which is named NanoBRET (*132, 133*).

The full-length protein of interest is transiently transfected in HEK293T cells with the mentioned Nluc at the C terminus. The luciferase Nluc is a 19 kDa protein that uses furimazine as substrate to emit luminescence at 450 nm (*134*), serving as BRET donor of the experiment. Also, a promiscuous inhibitor of the protein of interest, called tracer, is labelled with the BRET acceptor and incubated with the cells. In absence of an additional inhibitor, the tracer binds to the binding pocket in close proximity to the luciferase and absorbs its light through an energy transfer, analogue to the fluorescence resonance energy transfer (FRET). The acceptor emission is measured at 650 nm. Upon addition of the inhibitor in question, the tracer is displaced from the binding pocket, leading to a loss of energy transfer and therefore a higher percentage in donor emission. To calculate the BRET ratio, the acceptor/donor emission ratio from the donor only is subtracted from the acceptor/donor emission ratio of the sample:

$$(7)\ BRET\ ratio = \frac{Acceptor\ channel\ emission\ of\ sample}{Donor\ channel\ emission\ of\ sample} - \frac{Acceptor\ channel\ emission\ of\ donor\ only}{Donor\ channel\ emission\ of\ donor\ only}$$

mBRET units are calculated by simply multiplying the BRET ratio by 1000:

$$(8)\ mBRET\ unit = BRET\ ratio * 1000$$

The subtraction of the acceptor/donor emission of the donor only control is necessary, because the raw acceptor channel emission consists of not only energy transfer emission, but also overlapping direct donor emission.

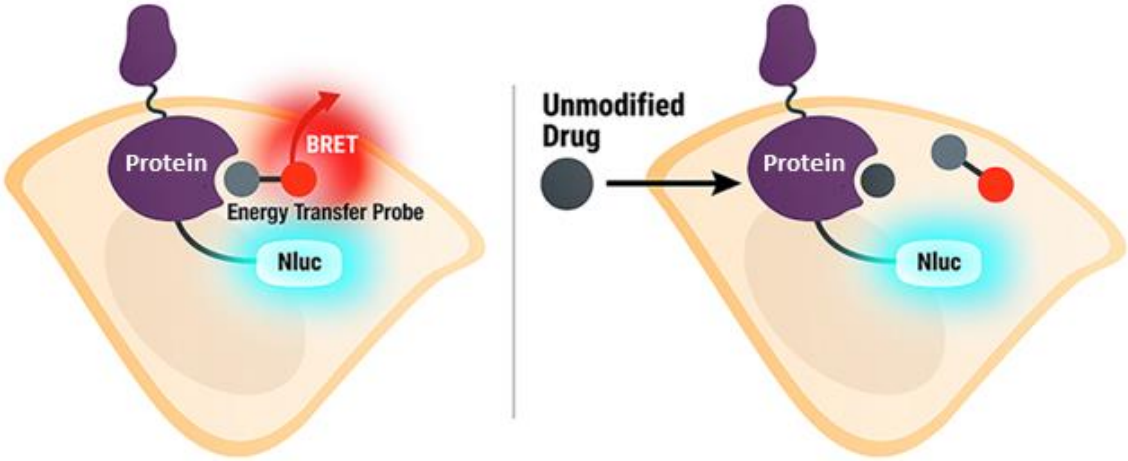


Figure 12 Scheme describing the principle of a NanoBRET experiment. In the absence of sample (unmodified drug), the luciferase on the protein of interest transfers energy to the acceptor-labelled tracer, leading to acceptor emission (left). Interaction of the sample interferes with the protein-tracer interaction and donor emission increases (right) (adapted from (*133*), CC BY).

Protocol NanoBRET

The full-length ALK and BRD4(1) plasmid containing C-terminal placements of NanoLuc were obtained by the manufacturer (Promega). To lower intracellular expression levels of the reporter fusion, the NanoLuc/kinase fusion construct was diluted into carrier DNA (pGEM3ZF-, Promega) at a mass ratio of 1:10 (mass/mass), prior to forming FuGENE HD complexes according to the manufacturer's instructions (Promega). DNA:FuGENE complexes were formed at a ratio of 1:3 (µg DNA/µl FuGENE). 1 part of the transfection complexes was then mixed with 20 parts (v/v) of HEK293T cells suspended at a density of $2x10^5$/ml in Dulbecco's modified Eagle's medium (DMEM) (Gibco) + 10% FBS (GE Healthcare), seeded into T75 flasks and allowed to express for 20 h. For target engagement both serially diluted test compound and NanoBRET Kinase Tracer K5 (ALK) and BRD-Tracer (BRD4(1)) (Promega) at a final concentration of 2µM and 0.5 µM, respectively, were pipetted into white 96-well plates (Corning 3600). The corresponding ALK or BRD4BD1-transfected cells were added and reseeded at a density of $2x10^5$/ml after trypsinisation and resuspending in Opti-MEM without phenol red (Life Technologies). The system was allowed to equilibrate for 2 h/37°C/5% CO_2 prior to BRET measurements. To measure BRET, NanoBRET NanoGlo Substrate + Extracellular NanoLuc Inhibitor (Promega) was added as per the manufacturer's protocol, and filtered luminescence was measured on a CLARIOstar plate reader (BMG Labtech) equipped with a 450 nm BP filter (donor) and a 610 nm LP filter (acceptor). Competitive displacement data were then graphed using GraphPad Prism 7 software using a 4-parameter curve fit with equation 9:

$$(9)\ Y = \text{Bottom} + \frac{(\text{Top}-\text{Bottom})}{\left(1+10^{(\log IC_{50}-X)*\text{HillSlope}}\right)}$$

X-ray crystallography

The way from a first hit to an affine inhibitor can often be a long one, especially because not only potency, but also specificity and physicochemical properties have to be considered. Certainly, modern equipment enables research to screen many thousand inhibitors per day, making the identification of potent inhibitors fast, but also costly. Furthermore, high-throughput screening often does not deliver much information about how a potent inhibitor can be further improved or how its specificity can be adjusted to avoid certain off-target effects. In this aspect, combination of biochemical and biophysical profiling of inhibitors with structural models of the protein-inhibitor complex facilitates a more rational way to understand and improve the desired interaction.

At the end of 2018, the protein data bank (PDB) contained 123,167 protein structures solved by X-ray crystallography, 10,934 protein structures solved by nuclear magnetic resonance (NMR) spectroscopy and 1,939 structures by electron microscopy (*67*), showing that X-ray crystallography was the method of choice for over 90% of all protein structures. This is particularly relevant from a drug design point of view, because structural data generated with NMR is often restricted to either small proteins (smaller than 10 kDa), few residues in larger proteins and often expensive labelling techniques. Electron microscopy, on the other hand, being a powerful method to gain information about quaternary structures of proteins, still lacks the ability to resolve protein-inhibitor interactions in most cases. Although especially electron microscopy is evolving at fast pace, X-ray crystallography is still the most dominant method used for obtaining information about protein-inhibitor interactions on a molecular resolution.

The basis for using X-rays as a means to create a model structure for a protein of interest is their wavelength, which is in the range of atomic bonds (~0.15 nm). Although a real macromolecular X-ray dataset seldom resolves hydrogen bonds, this is the theoretical limit of such data. To create a homogeneous set of planes for diffraction, protein molecules are crystallised into repeating conformations and orientations, before they are subjected to X-ray radiation.[1]

[1] If not noted otherwise, the content in this chapter is derived from (*135*).

Sitting drop vapour diffusion

Inorganic salt crystals can be generated by heating a saturated solution and slowly cooling it down, while a thread can be used as nucleation point for the crystal. Alternatively, organic solvent can be slowly added to the solution to induce crystallisation. Both methods base on the change of the solubility product up to a point, where the concentration of compound is too high to remain soluble. Although heating and organic solvent do not apply for the crystallisation of proteins, this process can be mimicked with the addition of a so-called precipitants, e.g. ammonium sulphate or polyethylene glycol (PEG). Salts promote intermolecular interactions by depleting water from the soluble hall of the protein at high concentrations, whereas polymers like PEG promote volume exclusion effects and therefore change solubility in the microenvironment of the protein.

A phase diagram with three major phases describes the process of crystallising protein (Figure 13). Protein concentration and precipitant are the determinants that define whether the protein precipitates, whether it is supersaturated or in a stable, soluble state. In an optimal crystallisation experiment, nucleation occurs in supersaturated conditions ("Labile") and few crystals grow in size in the metastable area of the phase diagram. Practically, this is achieved by vapour diffusion between a drop and a reservoir of much larger volume than the drop. A setup, where the drop is pipetted onto a plate is referred to as "sitting drop", whereas the attachment to an inverted cover slip is referred to as "hanging drop". The drop contains a mixture of purified protein and precipitant solution, resulting in an undersaturated condition. This solution equilibrates with the reservoir solution containing precipitant concentration optimal for crystal growth. Once the protein and precipitant concentrations reach critical concentrations, small crystals nucleate and reduce the surrounding protein concentration ideally to a metastable level promoting crystal growth (*136*).

If, for a specific case, the labile phase is so thin that nucleation is unlikely to occur, micro seeds of existing crystals can serve as attachment surface for protein molecules. On the other hand, if the metastable phase is difficult to hit in a specific case, nucleation will continue to occur, and one is likely to obtain many small crystals. In order to find suitable precipitant conditions, so-called coarse screens are used in first instance as means to screen for different precipitants in combination with different pH values and salt contents. Suitable protein concentrations can be estimated by eye judgement immediately after pipetting a crystallisation plate: if roughly half of the drops contain precipitate and half of the drops are clear, one is likely to have appropriate concentrations for at least some conditions. In some cases, the protein cannot be crystallised on its own, even if nucleation is principally facilitated by microseeding. Sometimes, the addition of a ligand enables crystallisation by either restricting flexible

regions of the protein to fewer number of possible conformations or by changing the conformation of the protein to one that is more suitable to interact in repetitive ways to form a crystal.

Crystal growth is observed in short intervals in the first days after plate setup and in longer intervals during week and months, which is due to the kinetics of the changing equilibria. Once crystals are obtained, the best way to judge their experimental qualities is to test the diffraction upon X-ray exposure. However, since beam time or accessibility to a local radiation source usually is restricted, the most promising crystals are identified visually. Optical clarity, sharp edges and smooth surfaces in single crystals are reasonable identifiers for diffracting crystals. Additionally, UV absorption and sharp changes between brightening and darkening under rotating polarised light can be consulted. Suitable crystals are subjected to an optimised condition with a cryoprotectant like PEG or glycerol, mounted and flash-frozen in liquid nitrogen. All steps from there are performed under 100 K or lower, reducing radiation damage and X-ray scattering from water molecules during data collection and enabling long-term storage of crystals. The use of cryoconditions is commonly referred to as cryocrystallography.

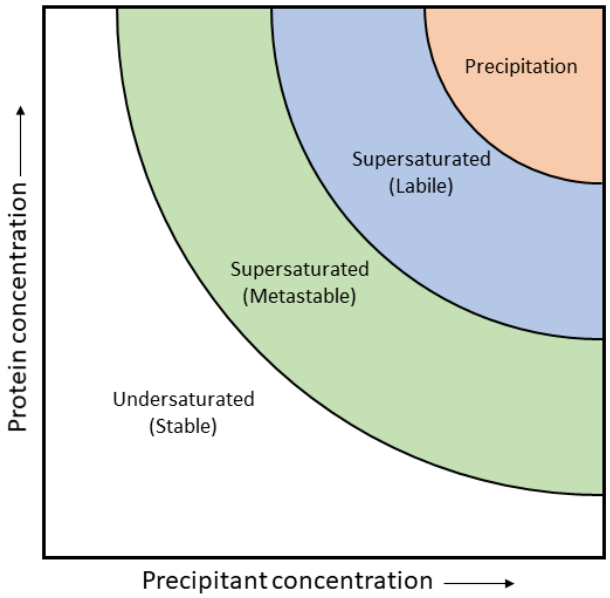


Figure 13 Phase diagram for nucleation and growth of protein crystals. The mixture of protein and-precipitant gradually must reach a supersaturated state to form crystals. Too high concentrations lead to precipitation and too low concentrations favour the soluble state of the protein

Protocol Siting drop vapour diffusion

Swiss CI 96 well 3 lens low profile plates are used for sitting drop vapour diffusion. First, 20 µl of precipitant solution is pipetted into the reservoirs by hand. Compositions of the used screens are listed in Supplementary Tables S2-S6. A mosquito pipetting robot then distributes protein from a source plate to the three drop positions of each well in the crystallisation plate and adds reservoir solution in different ratios. For each condition, protein:precipitant ratios of 130:70 nl, 100:100 nl and 70:130 nl to explore different parts of the crystallisation phase diagram. After pipetting, the plate is immediately sealed with qPCR seal and drops are observed under a microscope to judge the drop quality and the degree of precipitation. The plates are stored in incubators at either 277 K or 293 K and observed under a microscope after one day, three days, one week, two weeks and one month. For transport and storage, 0.5 µl mother liquor containing 25% ethylene glycol is added to the drop, the crystal is mounted onto a loop and transferred into a Unipuck submerged in liquid nitrogen.

Geometric basis for diffraction

When subjecting protein crystals to X-rays, one can observe distinct diffraction spots if the emerging reflections are collected on a photographic film or a detector. In the beginning of the 20th century, William Lawrence Bragg discovered that these reflections, emanating at different angles from the crystal, can be computed, if they are treated as if they were reflection from sets of equivalent, parallel planes of atoms in the crystal.

One aspect of describing crystallographic planes is the definition of a so-called unit cell with edges **a**, **b**, and **c** of lengths *a*, *b* and *c* and corresponding angles α, β and γ. A unit cell is the smallest assortment of molecules with defined dimensions and angles, with which the whole crystal can be created by simply stacking together these parts. There are seven unique arrangements called crystal systems based on equalities in cell edge lengths and angles, which build the basis for thirteen unique lattice types (Table 12).

Table 12 Three-dimensional crystal systems

Crystal system	Edge lengths	Angles
Cubic	$a=b=c$	α=β=γ=90°
Tetragonal	$a=b\neq c$	α=β=γ=90°
Orthorhombic	$a\neq b\neq c$	α=β=γ=90°
Rhombohedral	$a=b=c$	α=β=γ≠90°
Hexagonal	$a=b=c$	α=β= 90°, γ=120°
Monoclinic	$a\neq b\neq c$	α=γ=90°, β≠90°
Triclinic	$a\neq b\neq c$	α≠β≠γ≠90°

The equalities of unit cell edges do not only describe same dimension but identify the internal symmetry of the unit cell. A unit cell with *a*=*b*=*c* and α=β=γ=90° is only clinic, if all contents along the three axes are equal. The basic crystal systems are complemented by three different lattice types: a primitive lattice with 8x(1/8) points per cell (P), a body-centred or internal lattice with an additional, thus 2 lattice points per cell (I) and a face-centred lattice with additional points on the faces of the cell (F). The subdivision of the seven crystal systems in the three lattice types gives thirteen unique lattice types, called *Bravais* lattices. By definition, a more symmetrical division of lattice points into a unit cell is favoured over a less symmetrical choice.

Space groups describe the afore-mentioned lattice type and internal symmetry present in a unit cell. In total, it is possible to construct 230 space groups. For instance, simple symmetry operations are translation, rotation along an axis or reflection, which can be illustrated by duplication along a mirror plane. In fact, protein molecules are chiral molecules and two proteins cannot be placed into a unit cell to generate mirror symmetry. For this reason, symmetry operations in protein crystallography are limited to translation, rotation and their combination. This reduces the number of possible space groups from 230 to factual 65 in the context of protein crystallography. Translation simply describes movement of a molecule along a vector, while rotation changes the orientation of the molecule relative to a fixed point. Often, a combined symmetry operation of translation and rotation named “screw axis” is observed. It is represented by the symbol n_m, a translation of m/n of the unit cell on an n-fold screw axis. For example, $P2_1$ is a primitive unit cell containing a two-fold screw axis parallel to c, thus molecules along that axis alternate in their orientation by 180°. A $P3_13_13_1$ contains three perpendicular screw axes parallel to the three edges. In this case, molecules divide one unit cell length into three equal parts, while rotating 120° per symmetry operation. Knowledge about the unit cell symmetry reduces the experimental effort, because e.g. in a simple 2-fold-symmetrical unit cell, data collection after more than 180° is theoretically redundant. However, in practice, redundant collection of equivalent information (called reflections) leads to more accurate data. For the assembly of a unit cell, more than one protein molecule per point may be necessary. The smallest set of molecules which can reproduce the whole unit cell with the described symmetry operations is called *asymmetric unit*.

Further description and assignment of crystallographic planes allows the determination of unit cell dimensions and gives a basis for the understanding of diffraction. The most basic planes are those forming the faces of a coordinate system, assuming that x aligns with **a**, y with **b** and z with **c** of the unit cell. So-called *Miller indices* h, k and l define a set of planes, which leads to a specific diffraction angle. Each index describes the number of planes along one axis of the coordinate system, or equivalently,

how often per unit cell that axis is cut by a specific set of planes. The planes (100) with h=1, k=0 and l=0 describe the surface that is spanned by axes y and z, making them cut axis x once per unit cell dimension. Planes (123) cut x once per unit cell, y twice and z three times, whereas all (123) are parallel and have an equal spacing d between each other. A vector perpendicular to one plane and its sign in each direction determines the signs of each indices h, k and l. It does not matter which of the two perpendicular directions this vector faces to, as this will result in opposite indices for all three, being completely equivalent.

Bragg discovered that a set of parallel planes reflects an X-ray beam with wavelength λ and impinging angle Θ at the same angle to produce a new coherent beam, if the condition

$$2d\ sin\theta = n\lambda \quad (10)$$

is met with n as an integer number. By basic trigonometry, 2dsin Θ is the distance that a more deeply penetrating beam additionally travels to produce a reflection with same phase than the beam being reflected from a neighbouring plane (Figure 14).

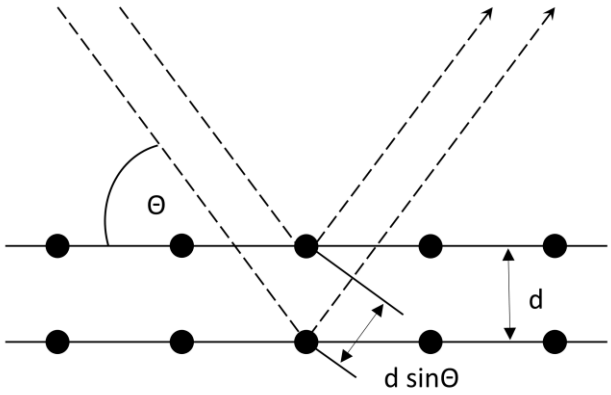


Figure 14 Geometric scheme showing Bragg's law. Only for a specific angle Θ, X-rays will be reflected coherently by a set of planes to produce a diffraction spot.

The equivalents to diffracting planes are electron clouds in the protein. It is the aim of crystallographic experiments is to obtain a graph of a mathematical function ρ(x,y,z), which represents the electron density for all coordinates in the unit cell. When collecting the diffracted beams on a film or detector, one can observe distinct spots from all crystallographic planes, each leading to another diffraction angle. Although only being obvious for diffraction patterns of simple objects, distances between coordinates in the real space (the crystal itself) and distances between correlating diffraction spots have an inverse

relationship. All reflections are therefore part of the so-called reciprocal space and are related to coordinates of the reciprocal lattice.

The reciprocal lattice is constructed by choosing an arbitrary point O as the origin of the real lattice and reciprocal lattice (Figure 15A). From a neighbouring point, a representative from each set of planes is drawn. By drawing lines orthogonal to the planes with the inverse interplanar distance 1/d of the respective set, the reciprocal lattice points are obtained. Bragg's law in the reciprocal space can be elucidated by a geometrical construction (Figure 15B). In its centre lies a crystal C with a crystallographic plane (red) in the real lattice. A circle is drawn around the crystal and an X-ray beam passes through point B and the crystal C to the origin of the reciprocal lattice O. Also, the beam is diffracted by the crystal with a total angle change of 2Θ. Point P and P* are points in the reciprocal lattice. By choosing 1/λ as the radius of the circle and connecting points B and P, one can visualise Bragg's law. According to Thales's theorem, angle BPO is rectangular and therefore

$$\sin(\theta) = \frac{OP}{OB} = \frac{OP}{\frac{2}{\lambda}} \qquad \text{and by rearranging} \qquad \frac{2}{(OP)}\sin(\theta) = \lambda \,. \tag{11}$$

Using the knowledge from the construction of reciprocal lattice points in Figure 15A, OP is the distance between O and a reciprocal lattice point and is equal to 1/d. Replacing OP with 1/d leads back to equation 10, which is Bragg's law. This construction therefore shows that only those reciprocal lattice points, which come in contact with the circle, yield diffraction. P* in Figure 15B is a reciprocal lattice point, but can only fulfil Bragg's law, if the reciprocal lattice is turned around O. In the real experiment, this is achieved by turning the crystal. Because diffraction takes place in three dimensions, the black circle in Figure 15B is actually a sphere named *Ewald sphere*. The projection of reciprocal lattice points on the *Ewald sphere* onto a two-dimensional plane as done with a photographic film or a detector yields the characteristic *loons* in a diffraction pattern.

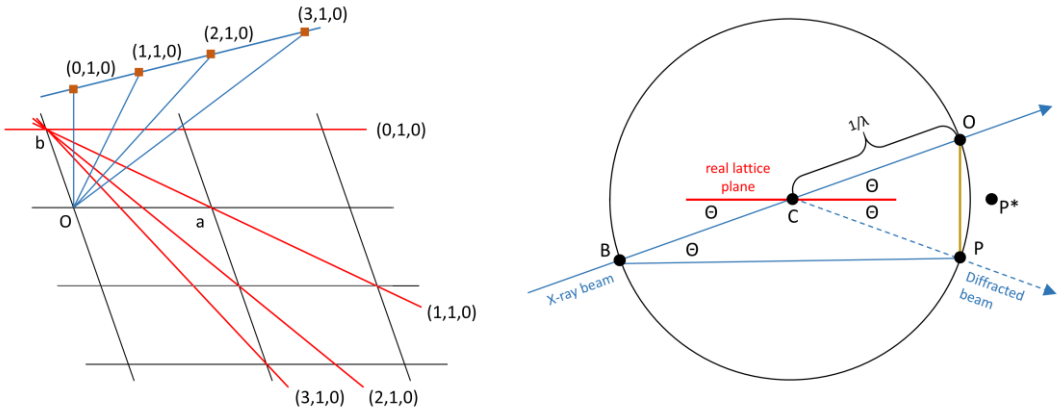


Figure 15 A Geometric construction of reciprocal lattice points **B** Geometric representation of Bragg's law. A reciprocal point P comes in contact with a circle of radius 1/λ, if a crystallographic plane at position C in the middle of the circle diffracts an X-ray beam at a suitable angle Θ.

Mathematical relationship between reflections and electron density

As discussed above, reflections in the reciprocal lattice are products of X-rays diffracted by crystallographic planes in real space, that is to say the crystal itself. The aim of the experiment, a three-dimensional map of electron density in real space, is non-trivial, since the mathematical description of each diffracted beam as waves with amplitude F, frequency h and phase α is as complex as the arrangement of atoms in a protein structure. However, Jean Baptiste Joseph Fourier showed that complex periodic functions can always be represented by an infinite sum of simple periodic functions. Furthermore, he could prove that for any function f(x), there exists another function F(h) such that

$$F(h) = \int_{-\infty}^{+\infty} f(x) e^{2\pi i(hx)}\, dx. \tag{12}$$

Function F(h) is called *Fourier transform*. It transforms units of x in reciprocal variables h, which is exactly what a crystallographer seeks, since the experimental information in reflections are in reciprocal relationship to electron clouds surrounding the protein atoms. Equation 12 is a one-dimensional function that can be translated into a three-dimensional relation.

$$F(h,k,l) = \int_x \int_y \int_z f(x,y,z) e^{2\pi i(hx+ky+lz)}\, dx\, dy\, dz \tag{13}$$

The basis for each reflection is its corresponding diffracted X-ray, whose mathematical description itself is a Fourier sum called *structure factor* F_{hkl}. There are two ways to interpret the contributions to a structure factor. One way describes a structure factor as a Fourier sum of atomic structure factors f_{hkl}, in which each atom in the unit cell contributes to the wave properties.

$$f_{hkl} = f_j e^{2\pi i(hx_j+ky_j+lz_j)} \quad (14)$$

$$F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hx_j+ky_j+lz_j)}, \quad (15)$$

where f_j represents a scattering factor that accounts for an atom to be treated as a simple sphere, x, y and z are the atom coordinates in real space and h, k and l are indices of a specific reflection in the reciprocal lattice. The indices h, k and l also describe the frequency of the wave. The contribution of each atom j in the cell depends on its element (because of different amounts of electrons), giving rise to a specific amplitude f_j, and also depends on its position in the unit cell. The other way to describe a structure factor is as sum of contributions from infinitesimally small volume elements of electron density in the unit cell.

$$F_{hkl} = \int_x \int_y \int_z \rho(x,y,z) e^{2\pi i(hx+ky+lz)}\, dx\, dy\, dz \quad (16)$$

Consequently, electron density on real (hkl) planes can be expressed as Fourier sum of structure factors F_{hkl}.

$$\rho(x,y,z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)} \quad (17)$$

Electron density is in fact a real space entity, although equation 17 contains an imaginary term in the basic wave functions. This is because for every positive index h, there is an equivalent negative index h, which cancels out the imaginary term, a concept that is implied in Friedel's law. However, the imaginary term is still relevant for the phase of the structure factor, being represented by the term $2\pi i(hx_j+ky_j+lz_j)$ in the exponent of equations 14-17.

As shown above, the amplitude of the complex wave function is represented by f_j, practically being proportional to the square root of the reflection intensity I_{hkl}, and the frequency is determined by indices h, k and l, which can be constructed with the help of the described geometric relations. However, the phase cannot be obtained directly from the diffraction pattern, an obstacle that is known as the *phase problem* in crystallography. In this thesis, this problem is approached by a computational approach called molecular replacement, being described in a later paragraph.

Data collection

In general, the collection of crystallographic data requires an X-ray beam source, a construction that holds, rotates and cools the crystal in the beam and a detector collecting the reflections. In the first instance, the aim of data collection is the determination of two parameters for each reflection: distinct

values for h, k and l and an intensity (darkness of spot) which depends on the amount of electrons contributing to the reflection.

An X-ray is characterised as electromagnetic wave with wavelengths between 0.1-100 Å. Basically, they can be produced by bombarding a piece of metal such as copper or chromium with electrons which in turn emerge from a heated filament. However, especially reflections from planes with short interplanar spacings will only be detectable, if the beam is narrow and intense. Therefore, particle storage rings at particle accelerators, in which electrons or positrons are accelerated to nearly light speed, are the facilities of choice. To achieve such speeds with limited space, the facilities are often ring-shaped, and the charged particles are forced onto a circular motion, a process which produces synchrotron radiation tangentially to the ring. A so-called *wiggler* bends and therefore intensifies the beam further, while mirrors and monochromators focus the beam into an experimental chamber outside of the ring.

The focused synchrotron beam in the experimental chamber is directed towards the mounted crystal, which is attached onto a so-called goniometer head. A goniometer head is an assembly of motors performing fine motions to on the one hand centre the crystal precisely into the beam position and on the other hand rotate it to move all possible points of the reciprocal lattice into contact with the Ewald sphere (cf. Figure 15). A nitrogen stream next to the goniometer head keeps the crystal at temperatures close to 100 K to prevent it from sustaining severe radiation damage, which would especially affect macromolecules subjected to radiation at room temperature. Most of the X-ray waves will pass through the crystal undiffracted. They are prevented to damage the sensor by a beam stop, a small round piece of metal between crystal and detector.

Nowadays electronical area detectors replaced conventional photographic film. Similar to digital camera devices, charged-coupled devices (CCD)-based or complementary metal-oxide-semiconductor (CMOS)-based solid-state devices are used, which accumulate electron charge proportional to the number of absorbed photons. Since X-rays are not in the spectrum of visible light, substances like phosphor are used as means to emit photons upon subjection to X-ray radiation.

Before starting the collection of data for a range of angles, suitability is assessed in a characterisation measurement. To do so, the crystal is mounted and centred into beam position, which works best at the “face-on” view displaying the loop on its flat side (*137*). When identification of the crystal position within the loop remains challenging due to crystal size/optical reasons, a grid scan with low energy radiation bursts at different positions of the loop can be a helpful method. Then, two pictures 90° apart are collected, because frequently diffraction differs at different orientations. Suitability for data

collection and collection parameters are judged based on observed maximal resolution, cleanliness of the reflections and estimation of the number of lattices.

If the decision falls for a full data collection, tools like iMOSFLM assist in the development of a collection strategy (*138*). The characterisation pictures are analysed regarding strong and weak spots to calculate possible unit cell dimensions and symmetries and Miller indices are assigned. This process is called *indexing*. After judging the correctness of low-resolution spots and predicted spots for the respective solution, the experimentalist chooses the solution with highest symmetry. As intermolecular interactions in protein crystals are weaker than in inorganic crystals, the arrangement of unit cells in a protein crystal rather resembles a mosaic than a perfectly aligned array. Diffracted X-rays emerging from the crystal will therefore be cone-shaped and spread over a small range of angles, a phenomenon called *mosaic spread*. Mosaicity should also be considered in the selection of an indexing solution. The program plots the total predicted intensity against mosaic spread and defines a value in degrees, where the intensity reaches a plateau. Usually, this plot is sigmoidal. Depending in the selected symmetry, the program advises the experimentalist with a suitable starting angle and range for data collection.

Additional parameters for data collection are dependent on the beam properties and estimated spot intensity regarding the radiation sensitivity of the crystal. If, for example, a relatively short exposure time during characterisation yields reflections with good intensities, it is advisable to keep exposure time and transmission at the lower limit to lower the risk for radiation damage. Also, consultation with the beam line scientist about how strong the beam is, and which method is used to focus the beam and adjust its size, is often helpful. The detector distance is decided based on observed resolution in the characterisation and cleanliness of spots. For instance, in case of overlapped spots, where long unit cell axes are oriented close to the beam, it is often advisable to sacrifice resolution for better spot separation. In addition, nowadays, an oscillation method is used. In this method, one frame is collected by oscillating the crystal about a fraction of an angle ϕ so that different reflections appear. Precessing the crystal to the next angle and using oscillation will lead to overlap of some already observed spots and the appearance of new spots. This method is useful to prevent spatial overlap and smaller oscillation angles should be considered in trickier cases. During and after data collection, data processing in XDSapp is a powerful tool to quickly assess data collection parameters with plots in a graphical user interface (GUI). For example, the number of strong spots against collected frames gives some indication of radiation damage.

Protocol Data collection

After mounting the crystal onto the goniometer head, the loop is centred under the highest possible zoom to ensure optimal orientation of the crystal. To assess the diffraction properties of the crystal, two frames are recorded from the "face-on" side and the 90° apart "side-on" view. The acquisition parameters are set to 0.5 s exposure time, 0.5° oscillation range, 2 Å (equ.) detector distance and 50% transmission. Diffraction pictures are judged with *adxv* and a data collection angles are obtained with iMOSFLM (*138*). Usually datasets are collected at 0.1-0.2° oscillation angle and exposure time and total number of frames are set depending on the symmetry and he beamline. Crystal integrity is observed with XDSapp which is also used to estimate the quality of the data and confirm space group and cell parameters (*139*).

Data processing

After all necessary reflections are collected at the synchrotron, data processing of the collected data either begins with the mentioned program iMOSFLM or XDS (*138, 140*). Both programs include indexing, calculation of unit cell parameters and integration of reflections. iMOSFLM has a general advantage of a GUI, giving the user graphical feedback for all relevant processes and therefore being more intuitive for unexperienced users. XDS is run over a script or directly from the command line and requires more background knowledge about the underlying processes.

Autoindexing is an automated process which determines crystal orientation, rough unit cell parameters and unit cell symmetry and assigns Miller indices to the reflections. Since the recorded images are collection of pixels to start with, the programs have to rationalise the pictures with regard to background counts, the intensity of spots over background, their distribution across the images, shape of the spots and number of pixels per spot (*141*). iMOSFLM uses spots on single images that are separated over ϕ (typically 90°). XDS, on the other hand, combines this approach with a 3D component by using profiles of two sets of consecutive sets of images, which are separated over ϕ angles. After the definition of spots against background is finished, two steps are important for indexing. In the first, the reflections are mapped in the reciprocal space by relating them to their respective positions in the Ewald sphere, defined by so-called scattering vectors. In the second step, the mapped reflections are related to different unit cells. This is either achieved with a real space method, where the scattering vectors are first transformed into real space before the search is performed, or with a reciprocal space method, where the search is performed first and then the vectors are transformed into real space.

Basically, all programs will list solutions with different unit cell parameters and symmetries with penalty scores, which are based on how well the experimentally found spots represent the particular unit cell.

Once a choice for a solution has been made by the experimentalist, the determined unit cell parameters must be refined with regard to imperfect crystal to detector distance, crystal slippage during the data collection and generally by considering a greater fraction of the data. Two strategies can be used: positional refinement or post-refinement (*141*). Positional refinement calculates discrepancies between experimentally determined and calculated spot positions. Post-refinement determines the relative intensities of partial reflections throughout many frames and therefore requires integration of data. iMOSFLM uses a combination of both methods – positional refinement for refinement of the crystal-detector distance and post-refinement for unit cell parameters. XDS refines the parameters after integration.

In the integration step, intensities are obtained for reflections. In iMOSFLM, the described spot profiling is expanded to the whole dataset and spot intensity is defined in two dimensions by the number of pixels per spot and darkness of the pixels compared to the background. XDS uses consecutive frames to generate 3D profiles of reflections. Integration parameters are set depending on the used detector as each detector differs slightly (signal to noise etc.).

When all reflections are collected, indexed and integrated, one can determine the space group with higher accuracy, which is performed by the program POINTLESS as part of AIMLESS (*142*). One underlying principle of determining the space group with higher accuracy are so-called *systematic absences*, reflections which are missing as implication of the unit cell symmetry. The phase of a structure factor is equal to $2\pi(hx_j+ky_j+lz_j)$ as exponential term in equations 14-16, if only relative phases of two equivalent atoms are considered. For a space group $P2_1$ with a screw axis along **b**, the phase contribution of an atom j at the origin (0, 0, 0) to the structure factor for plane (010) F_{010} is equal to $2\pi(0*0+1*0+0*0)=0$. Its symmetry-related atom j' lies at a relative position (0, +½, 0) and contributes to the same structure factor with phase $2\pi(0*0+1*1/2+0*0)=\pi$, being opposite to its counterpart and therefore cancelling out F_{010}. The same is valid for all odd-numbered 0k0 reflections. However, looking at the phase contributions to F_{020} of two equivalent atoms j and j' in the same $P2_1$ space group, they will contribute with 0 and 2π, being in phase and leading to a reflection. Another underlying principle in POINTLESS is the consistency of equivalent reflections. It analyses these conditions and outputs several solutions with corresponding probabilities.

Furthermore, AIMLESS performs *scaling* and *merging* processes. Reflections with same indices are ideally present multiple times in the dataset, but have inconsistent intensities, because not all angles of the crystal yield the same diffraction intensity, the X-ray path through the crystal has different lengths for different angles and the X-ray beam intensity itself slightly fluctuates during the data collection. In the scaling process, reflections with same indices are set to the same intensity, which is more accurate the higher the redundancy. Merging comprises two processes. First, it combines partial reflections with same index, which are distributed over consecutive frames. This is necessary in case of iMOSFLM, because profiling and integration both happen in two dimensions, but it is not necessary in XDS, because it is integrated in its 3D profiling approach. Second, merging equates Friedel pairs (reflection (h, k, l) is equal to (-h, -k, -l)) and reflections that are equivalent due to the space group symmetry.

Several parameters that are output by AIMLESS indicate how good the quality of the data is and gives the experimentalist rational means for *data reduction*, which means excluding of some part of the data. Finally, the program *truncate* transforms the integrated intensities into an estimated structure factor based on the probability distribution of the intensities.

There are several correlation factors which must be taken into consideration. First, there is R_{merge}, which calculates the spread of independent measurements of reflection intensities $I_j(hkl)$ around their average $I_{av}(hkl)$:

$$R_{merge} = \frac{\sum_{hkl}\sum_{j=1}^{n}\left|I_j(hkl)-I_{av}(hkl)\right|}{\sum_{hkl}\sum_{j=1}^{n} I_j(hkl)} \qquad (18)$$

Data at lower resolution is generally better distinguishable from noise and ideally should not exceed 8%. As mentioned in the integration description, redundant data improves the accuracy of intensities, because outliers have a lower share in the average values. An overall redundancy is therefore favoured. Since data collection itself is not the bottleneck of a synchrotron trip, one is often confronted with more data than necessary, leading much higher multiplicities, but also to an artificially high R_{merge}. The indicator R_{meas} has been suggested to correct for high multiplicities by adjusting R_{merge} by a factor of $\sqrt{\frac{n}{n-1}}$. R_{pim} also considers the expected precision of $I_{av}(hkl)$ by further multiplication with the factor $\frac{1}{\sqrt{n}}$. Another frequently used indicator is $CC_{1/2}$, a modified version of the Pearson correlation factor CC. The unmerged data is divided into two parts, each containing a random half of each unique reflection (*143*). Thus, the higher the correlation value $CC_{1/2}$, the better the data quality. A reasonable value for $CC_{1/2}$ is 50% for the high-resolution shell of the data. Furthermore, a signal to noise ratio I/σI of at least 2.0 and an overall completeness of at least 90% are considered ideal. A final decision on which resolution to

choose is best made with refined phases at later stages of data refinement by judging whether addition of data leads to addition of information, judged by observing R_{free} (explained in refinement paragraph) and the electron density.

Protocol Data processing

If the datasets were collected on a DECTRIS EIGER X 16M detector or DECTRIS PILATUS3 6M detector, XDS is used for autoindexing, cell refinement and integration. In case of the DECTRIS PILATUS 2M detector data, iMOSFLM was used for autoindexing, cell refinement and integration. After integration, AIMLESS is used to determine the space group with higher accuracy (POINTLESS), for scaling and merging and transformation of the integrated intensities into initial structure factors. Based on AIMLESS statistics including graphs plotting R_{merge} against frames and the Wilson plot, data is reduced to ensure that the best part of the dataset is used. The resolution cutoff is usually decided based on I/σI (>2.0), completeness of the data (>90%), multiplicity (>3.0) and $CC_{1/2}$ (>50% in highest resolution shell).

Molecular replacement and refinement – obtaining and refining phases

The amplitudes and frequencies of the observed structure factors themselves will not yield a meaningful structural model because of lacking phase information. It is necessary to obtain initial phases for the target molecule either through experimental approaches (e.g. using heavy metal atoms and their unique diffraction properties) or through a sufficiently similar model structure. The latter, being known as molecular replacement, is nowadays the most frequently used method due to the increasing number of structures being available.

The structure factor phases of a molecule depend on the atomic coordinates within the unit cell. Therefore, model structures are more likely to contribute meaningful phases if the tertiary structure is presumably similar to the target structure, which can be the case for proteins which are either functionally or evolutionarily related, e.g. different kinases, which (almost) all catalyse the transfer of a phosphate group from adenosine triphosphate (ATP) to protein side chains. The general idea is to probe the position and orientation of the model structure within the unit cell of the target structure, so that both molecules superimpose. Next, the structure factors of the model structure with the new coordinates are calculated and their phases are used as initial estimates for the target structure.

In principle, one could probe all atomic coordinates of the model that are possible by unit cell symmetry, calculate the corresponding structure factors F_{calc} and compare its amplitudes $|F_{calc}|$ with measured $|F_{obs}|$ obtained through reflection intensities. However, even with symmetry constrictions,

calculation of three coordinates for position and three angles for orientation simultaneously would lead to very long and tedious computer operations. Dividing the search into two parts, one where an optimal orientation is achieved (called rotation search) and one where an optimal position is reached (called translation search), greatly reduces the amount of calculations compared to a simultaneous search. However, despite reducing the computation efforts, the disadvantage of splitting the search is that in the rotational search one cannot calculate F_{calc} directly since the coordinates remain unknown variables. There are two approaches aiming at this problem, one being the use of a Patterson map and one being a statistical maximum likelihood method. A Patterson map is constructed by mapping all interatomic vectors of a molecule at an arbitrary origin, so that the model and target molecule Patterson maps are equivalent and independent of their position. This independence of the position is exactly what is useful for the rotation search. By changing the three rotation angles one after another and comparing the agreement between amplitudes $|F_{calc}|$ of the model Patterson map with the amplitudes $|F_{obs}|$ of the target Patterson map, one can find the orientation with the best correlation. This is measured by integrating both Patterson maps unit cells and multiplying them for each probed position. Where a peak is in the model integrated map, but not in the target integrated map, the product of the two will be zero. Solutions with similar orientations will therefore yield a maximum number of products unequal zero. The second method for the rotation search is maximum-likelihood and is applied by PHASER, the program that is used in this thesis (*144*). It answers the question what rotation and translation of the model molecule in the new unit cell are the most probable to lead to similar $|F_{calc}|$ and $|F_{obs}|$ values. The rationale behind this strategy is based on the assumption that the spread over average of independent value is a Gaussian distribution (*145*).

The best orientation can either be found through a probability approach or by probing positions in the translation search. Comparison between model and target can be measured in correlation as R factor, a definition that is also useful in the later refinement steps:

$$R = \frac{\sum_{hkl}\left||\mathrm{F_{obs}}(hkl)|-|\mathrm{F_{calc}}(hkl)|\right|}{\sum_{hkl}|\mathrm{F_{obs}}(hkl)|} \qquad (19)$$

Phases with either high log-likelihoods or low R values (typically 0.3-0.4 is sufficient) are used as initial estimates for the structure factor and finally yield electron density in real space that resembles the target molecule.

The aim of the subsequent refinement process is to minimise the differences between measured amplitudes from the native data set and amplitudes calculated from the current model. Even if the

model protein during molecular replacement and the target protein have the same sequence, the initial density will contain errors due to e.g. differences in crystal packing, binding pocket conformations and loop conformations. The program Coot allows overlaying electron density maps and atomic coordinates of the current model and provides the experimentalist with tools to adjust the model based on the electron density interpretation (*108*). Two useful representations are the $2F_o$-F_c contour map and the F_o-F_c difference map. The former contours the whole molecule and helps to detect wrong conformations, while the latter shows positive and negative peaks to identify missing elements or atoms whose positions are not verified by any $|F_{obs}|$. This map interpretation is alternated with the adjustment of the coordinates on basis of the interpretation, which is done by the program Refmac (*146*). Both programs are part of the CCP4 suite (*107*). After coordinate adjustment, Refmac calculates the agreement between model amplitudes $|F_{calc}|$ and measured amplitudes $|F_{obs}|$ with equation 19, in this case being referred to as R_{work}. A similar parameter R_{free} is calculated, which uses an unbiased set of reflections (usually 5%) for $|F_{obs}|$ that have not been included in the data processing. Although R_{work} usually is lower than R_{free} by a few percent, the difference between both gives a good indication whether the model is overinterpreted. A rule of thumb for a well-refined model is the resolution cutoff divided by a factor 10 (R_{free} of 0.23 for a 2.3 Å dataset). The so-called *figure of merit* (FOM) is a statistical means to weight Fourier terms that have been added during refinement with a measure of phase quality. It is especially useful during the first steps of refinement. An errorless phase would lead to an FOM of 1.0 and it usually ranges from 0.75 to 0.95 for a refined model.

There are several refining methods to ensure the determination of phases close to the global minimum of R_{free}. First, it is important to remove as much model bias as possible in the first refinement steps. For datasets with decent resolution (better than ~2.5 Å) it is preferable to only use the peptide backbone of the model without sidechains. Adding the sidechains to the unbiased difference map decreases the probability to refine a local minimum in R_{free}. At lower resolution it is at least advisable to remove questionable loop conformations and obviously faulty side chains conformations. Step-by-step adjustment of the atomic coordinates by Refmac adds phase terms to F_{calc} and reveals further details in the electron density. After adding all amino acid side chains that show density (some might be flexible and therefore “invisible” in the difference map), water and possibly ligands are added stepwise. However, water molecules should be added with caution, since molecules with no possible hydrogen bond partner or a greater distance than 3.5 Å to the macromolecule are likely to be an interpretation of noise. Two other refinement methods worth noting are solvent flattening and non-crystallographic symmetry (NCS) refinement. Solvent flattening adjusts electron density between proteins based on the

fact that solvent regions of the unit cell should have constant intensities (*147*). NCS describes the presence of identical molecules in the unit cell which cannot be superimposed by symmetry operations of the present space group. Consequently, their intensities have not been merged during data processing and thus their structure factors are less distinguishable from noise. Especially at lower resolution and symmetric multimeric protein assemblies, their addition to one set of intensities can be exceptionally helpful in obtaining better phases.

Beside careful interpretation of the experimental data, knowledge about protein structures and the use of statistics of published structures aid the decision process during refinement. As described by G.N. Ramachandran, a protein backbone adopts only a limited number of possible configurations defined by their dihedral angles ψ and ϕ (*148*). Observation of the well-known Ramachandran plot during refinement is therefore essential. Furthermore, accumulation of data through published structures allows the definition of preferred side chain rotamers and supplements basic knowledge regarding bond lengths and bond angles. Decisions in favour of outliers should only be made if the electron density is unambiguous. These factors should be analysed with great care before the deposition of a structure in a validation process. MolProbity and the quality control server from the Joint Center for Structural Genomics (JCSG) are a powerful tools for validation which allow an analysis of the mentioned geometric restraints in combination with water molecule analysis, electron density analysis, clash ratings and many more (*149*).

Protocol Phasing and refinement (example)

Phaser is used for molecular replacement (*144*). PDB code 2OSS (*6*) serves as template for BRD4 structures, whereas PDB code 5J9S (*19*) serves as template for initial ENL YEATS structures. Correctness of the MR solution is judged based on log-likelihood gain, Z-score (usually >7.0) and observation of the output electron density map (presence of clashes). If the data resolution is ~2.5 Å or higher, the electron density map is refined with a C_α chain (side chains removed with *pdbset*) in Refmac5 to remove side chain bias and subsequently modified with *parrot* to facilitate autobuilding with *Buccaneer (107)*. After 3-4 cycles of refinement in Refmac5 (*146*) and model adjustment in COOT (*108*), waters are automatically added and another refinement step is added. After deleting incorrectly input waters and adding missing ones, the structures is refined once again. In case of structures containing inhibitor electron density, molecular *cif* files are obtained from Grade Web Server, imported into COOT and modelled into the structure. TLS refinement parameters are obtained from TLS server (*150*). Structure validation is performed with MolProbity (*149*).

Results

Development and characterisation of new bromodomain inhibitors for subfamily VIII

The first part of the thesis aimed at the development of inhibitors for BRDs. To date, 32 out of 61 human BRDs are untargeted and their potential as drug targets remains elusive (*12*). In combined efforts of the computational knowledge in Emmanuel Mikros' lab from the University of Athens and our biophysical method repertoire in Stefan Knapp's labs (Oxford/Frankfurt), we designed a strategy to cost-effectively identify hits with unique chemical scaffolds binding to so far untargeted BRDs. This project resulted in a µM cell-active hit for the fifth BRD of PB1 and was published in J. Med. Chem. on 12 September 2016 (*13*).

The binding pockets of the BRD family VIII have been predicted to have intermediate to difficult druggability and so far, the only potent chemical probe is PFI-3, being selective for PB1(5) and SMARCA2/4 (*26, 89*). The BRDs in this family reside in SWI/SNF chromatin remodelling complexes. All SWI/SNF complexes contain a central helicase and SMARCA2/4, while the polybromo-associated BRG1 factor (PBAF) complex also contains PB1, which comprises six distinct BRDs. SMARCA2/4 and PB1 are frequently mutated in cancers like epithelioid sarcoma and malignant rhabdoid tumours, potentially making them drug targets (*13*).

The strategy to find a novel hit in this project was divided into three phases. In the first phase, 260,000 compounds from National Cancer Institute/Developmental Therapeutics Program (NCI/DTP) repository were pre-screened by three different computational approaches. A set of 40 compounds with best ratings were then subjected to biophysical measurements in the second phase. After decision on the most interesting hit, a computational hydration analysis was performed in the third phase to assess the most promising modification possibilities for affinity enhancement (performed by V. Myrianthopoulos and colleagues from the Mikros lab, Figure 16A). The biophysical validation of the modified compounds was performed by me and M. Wanior.

The first phase used three orthogonal virtual screening approaches, a 2D similarity analysis, a molecular docking scoring analysis and a 3D similarity comparison. To perform the 2D similarity search, compounds in the repository had to be converted to fingerprint strings by Canvas (Schroedinger Inc.) and compared to the fingerprint of JQ1. Generally, the binding modes of BET inhibitors differ from family VIII binders, as shown later with PFI-3 after this study had started. PFI-3 protrudes deeper into

the BRD pocket, depleting four conserved water molecules (Supplementary Figure S3). However, JQ1 was the only potent BRD inhibitor, when this study had started before 2016. Although BRD4 is part of BRD family II (BET), it shares the same conserved tertiary structure with PB1 and has a sequence similarity of 61.6%. Thus, the hypothesis was that using JQ1 and BRD4 as templates allowed the identification of novel compounds that generally suited BRD binding pockets, if the parameters were chosen carefully. In this respect, all benzodiazepines (BZDs) from the NCI repository were excluded because they would bias the results towards JQ1 analogues. In order to choose the most suitable parameters for fingerprint derivation, a number of other known BRD binders for BRD4, CREBBP and BAZ2B were dispersed into a randomly selected subset from the NCI/DTP repository. The fingerprint derivation parameters were then rated on basis of whether the known binders would appear as hits in this test subset. The best sets of parameters were then used to create the fingerprint strings of the whole set of repository compounds without the known ligands and BZDs. Then, the Tanimoto distance, a statistical means to assess similarity between two sets, is calculated to yield hit compounds similar to JQ1. For clarity reasons, I will refer to this hit set as “set 1”.

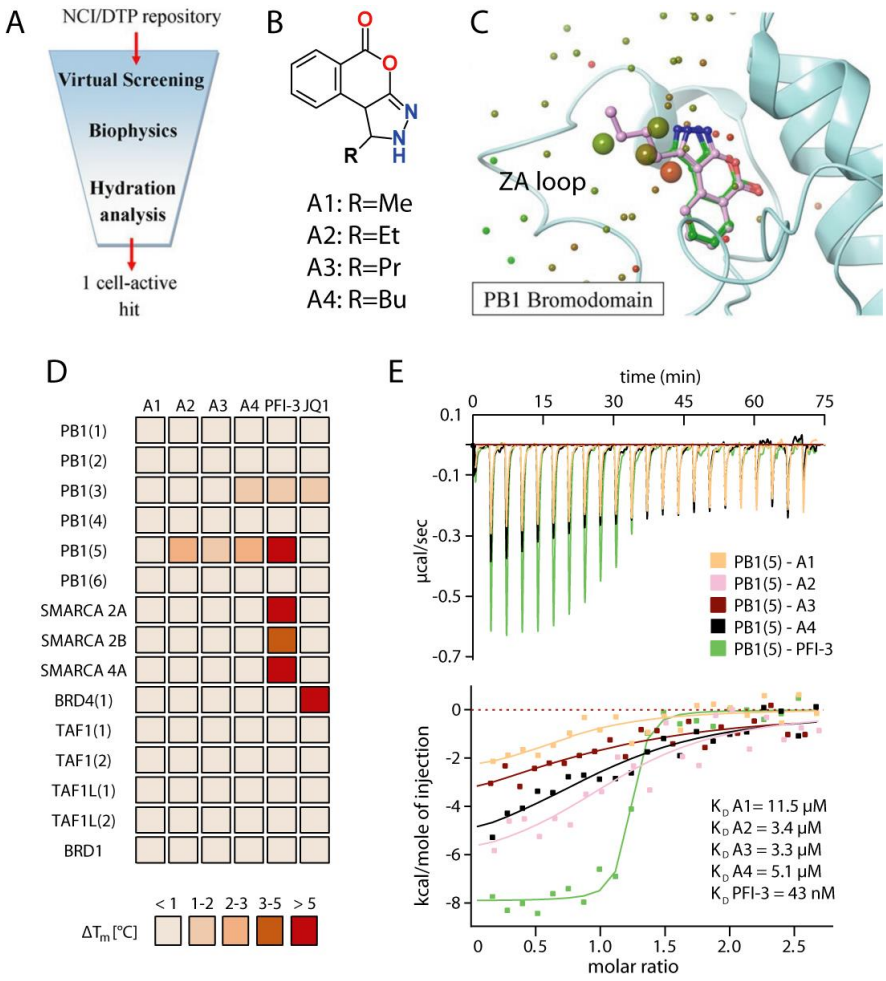


Figure 16 A Screening and optimisation strategy. Compounds from the NCI/DTP were preselected by three orthogonal virtual screening methods, characterised by biophysical methods and optimised on basis of a computational hydration analysis to yield a µM, cell active hit for PB1(5) **B** Hit compound A1 is a pyrazoloisocoumarin derivative modified to A2-A4 by elongating the alkyl chain **C** Results of the hydration analysis. Unstructured water molecules in a hydrophobic pocket are supposedly exchangeable by addition of and alkyl residue towards the ZA loop. **D** Thermal shift assay (100 µM compound concentration) comparing A1-A4 with reference compounds PFI-3 and JQ1. A1-A4 are selective for BRD family VIII. **E** ITC analyses of compounds A1-A4 binding to PB1(5). A 3.5-fold improvement in affinity is achieved by elongation of the alkyl chain comparing A1 and A3. (Experiments and Figures in A and C made by V. Myrianthopoulos, synthesis of molecules in A2-A4 performed by members of the Mikros lab (*13*))

5000 compounds with the best similarity rating from set 1 were further used for a docking scoring analysis with the GlideScore empirical scoring function as implemented in Glide. The scores served as basis to create a ranked list called set 2. Finally, a 3D similarity search was performed on all compounds in the repository. In contrast to 2D similarity, which relates the physicochemical nature of moieties in the compounds to molecular descriptors, 3D similarity covers overall molecular shape overlap (shape similarity), spatial similarity of pharmacophoric sites as well as positive and negative atoms, hydrogen bond donors and acceptors, hydrophobic moieties, and rings (colour similarity). These aspects were related to different conformers of the compounds in the repository, compared to the conformation of JQ1 in the binding pocket and then rated to form set 3 (*13*).

The virtual screening step was completed by comparing compounds in the top-ranking subsets of sets 1-3 in a consensus scoring approach based on their frequency of appearance. The 40 best-ranking compounds were identified as hits and subjected to biophysical analysis in phase 2 of the project (Figure 16A). The first step of phase 2 was to screen the 40 hits at 100 µM compound concentration against six BRDs including PB1(5) and BRD4(2) in a thermal shift assay. Among some hits with small Tm shift and one hit that resembled JQ1, compound A1 emerged as the most interesting molecule (Figure 16B). It stabilised PB1(5) by 4.1° C and BRD4(2) by only 1.9° C and was therefore further characterised by ITC and co-crystallisation with this domain. The K_D obtained from the ITC experiment amounted to 11.5 µM. The crystal structure was solved at 2 Å resolution and revealed a unique binding mode (Figure 16C, green molecule). The benzopyrone moiety was buried deep into the binding pocket and the N3 of the pyrazole ring formed an interaction with N739, equivalent to N140 in BRD4. The rest of compound A1 was mainly stabilised by hydrophobic interaction to L687 and L693 of the ZA loop, to I745 of αC helix and to I683 and F684 (which relate to PF in the WPF motif in BRD4).

Interestingly, A1 depleted four of five conserved water molecules from the binding pocket of PB1(5). BRD4 features conserved water molecules at similar positions, but all of them retain their position even if bound to JQ1 or other ligands. The depletion of water molecules was also confirmed by two other compounds (5,3',4'-trihydroxyflavone and luteolin) in the paper as well as salicylic acid and PFI-3 in a study published previously (*89*). These findings indicated that one possibility of targeting new binding pockets is the analysis of structured waters and their specific depletion by a ligand.

Intrigued by this exceptional binding mode, the aim of phase 3 was to enhance the affinity of A1 with chemical modifications. However, no moiety that could either form a hydrogen bond or an electrostatic interaction was evident in the vicinity of the compound. Therefore, we investigated if we can gain further knowledge about the hydration pattern to optimise the ligand. Three regions in the binding pocket were postulated to enhance the affinity of the inhibitor by depleting water molecules with modifications on A1: (1) the inner side of the acetyl-lysine cavity, (2) the entrance of the binding cavity, and (3) the periphery of the ZA channel. In this study, the hydration pattern in PB1(5) was analysed by the two programs SZmap (OpenEye Inc.) and WaterMap (Schrödinger Inc.). Region 1, the inner side of the binding cavity, coincided with the conserved water molecule that was not depleted by A1. Both SZmap and WaterMap predicted a coordinated hydration site at the same position. This water molecule is stabilised by interactions with side chain hydroxyl group of Y696, the side chain amide group of N734 and the backbone carbonyls of M704 and M731. Also for region 2, the entrance of the cavity, both programs predicted a structured water molecule. Thus, regions 1 and 2 could only be compensated by a polar group in the molecule that extensively forms the same hydrogen bond interactions with its environment as the water. As for region 3, the periphery of the ZA channel, the programs were not in complete agreement about the hydration state, but both predicted a rather unstructured network of many water molecules, presumably due to a rather hydrophobic sub-cavity.

Depleting these water molecules from region 3 by replacing them with hydrophobic modifications on A1 was chosen as strategy to enhance the inhibitor affinity. The moiety that pointed to periphery of the ZA channel was the methyl group in A1, depicted as R in Figure 16B. It was decided to stepwise elongate the methyl group to an alkyl chain, by adding one carbon atom at a time. This resulted in molecules A2 with an ethyl group, A3 with an *n*-propyl chain and A4 with an *n*-butyl group. Molecule A4, as representative with the longest alkyl chain, is superimposed in Figure 16C (pink molecule), showing how the network of unfavourable bound waters is targeted in this depletion strategy.

A1 and its analogues A2-A4 were synthesised and analyses with respect to their biophysical properties by a thermal shift assay and ITC experiments (Figure 16D and 16E). All six PB1 BRDs, SMARCA2, SMARCA4, the TAF BRDs and BRD1 were included in the experiment. PFI-3 and JQ1 served as positive references for BRD families VIII and II respectively. The thermal shift assay showed that all four compounds A1-A4 were selective towards PB1(5) and displayed higher thermal shifts than A1. A2 shifted the melting point the most and A4 the second most with a slight shift on PB1(3). A3 led to the least pronounced thermal shift among the modified compound A2-A4 but did not show off-target activity. The ITC experiment in Figure 16E confirmed that A2-A4 bind more effectively to PB1(5). A2 and A3 affinity improved by a factor 3.5 and A4 is double as affine as A1. The integrated heats show that the compounds produce considerable heats compared to the 43 nM binder PFI-3. Compound A4 was also tested with respect to cell activity in a fluorescence recovery after photobleaching (FRAP) experiment. Cells with green fluorescent protein (GFP)-tagged PB1 were treated with suberoylanilide hydroxamic acid (SAHA) to increase histone acetylation levels and therefore the assay window and incubated with different concentrations of A4. Then, FRAP half times $t_{1/2}$ were measured. DMSO served as negative control and PFI-3 as positive control. 20 µM A4 successfully reduced the half recovery time to a similar extent as PFI-3 at 1 µM which roughly corresponds with their affinity ratio towards PB1(5). Compound A4 and PFI-3 showed similar reduction of cell viability over 5 µM in 1BR-hTERT human fibroblast cells. However, incubation with A4 did not lead to sensitisation to DNA damaging agents like mitomycin C.

The study confirmed that not only the more obvious protein-ligand interactions but also hydration patterns effect the affinity of an inhibitor. A complementary approach of experimental methods like X-ray crystallography, thermal shift assay, ITC and computational methods like 2D/3D virtual screening and solvation mapping proved successful to identify a hit for BRD family VIII member PB1(5) and improve its affinity by chemical modifications.

Rational design of an ALK-BRD4 dual inhibitor

The second aim of this thesis was the design of novel kinase-BRD inhibitors. Ciceri *et. al* showed in 2014 that many potent kinase inhibitors also unintentionally inhibit BRD4 (*14*). BRD4 is a regulator of the oncogene *MYC* and therefore relevant to several cancer types (*151*). This dual inhibition effect is one explanation why adaptive resistance through compensating cellular pathways is less likely to occur for these inhibitors. The fact that different kinases shared the unintentional dual inhibition effect indicated that potentially other inhibitors could be rationally designed. Furthermore, many approved kinase inhibitors show some degree of polypharmacology which emphasises the potential of dual inhibitors

and also sets the function of chemical probes apart from inhibitors designed for a specific therapeutic effect (*11, 57*).

The following project was prompted by the mentioned study by Ciceri *et. al* and describes the development of an ALK^{F1174L}-BRD4 dual inhibitor as potential therapeutic agent for neuroblastoma patients on basis of BI-2536 (Figure 17). It was a collaborative effort between Ellen Watts, Swen Hoelder *et. al.* from the ICR in London, who were responsible for the chemical synthesis, docking analyses and the *in vitro* characterisation of inhibitors on protein kinases in particular ALK and the off-target PLK1, Monika Raab and Klaus Strebhardt from the University medical centre in Frankfurt am Main, who contributed cellular phenotypical assays on the cellular PLK activity of the developed inhibitors. In this collaborative effort, I performed structural biology analyses of interactions between BRD4/ALK^{F1174L} and the synthesised inhibitors by X-ray crystallography, *in vitro* characterisation on BRD4 by thermal shift assay and ITC and cellular target engagement assays towards ALK and BRD4 by NanoBRET. This work was accepted as article J. Med. Chem. in February 2019 (*16*).

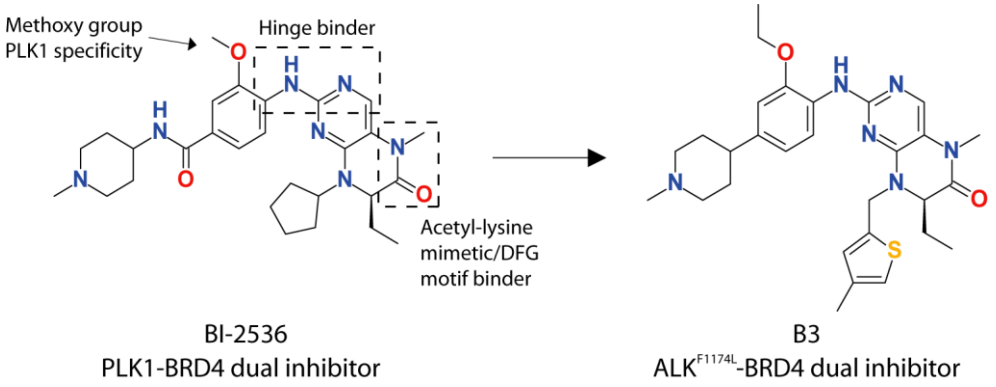


Figure 17 Development of an ALK^{F1174L}-BRD4 dual inhibitor. On basis of BI-2536, a dual inhibitor for PLK1 and BRD4, the target specificity was changed to ALK^{F1174L} and BRD4 with computational methods and the establishment of structure-activity relationships (SAR) with the help of biophysical methods (molecule B3 was synthesised by E. Watts)

Neuroblastoma is a cancer of neural crest origin occurring in childhood (*105*). High-risk patients harbour a mutated form of ALK (ALK^{F1174L}) that coincides with upregulated *MYC* gene activity (*106*). Although there are some selective ALK inhibitors in available or in clinical trials, many proved to be ineffective in that case presumably due to insufficient inhibition of the mutated kinase domain (*152*). Crizotinib and Ceritinib are examples for ALK WT-targeting drugs, whereas the recently developed lorlatinib is one of few compounds with high affinity towards ALK^{F1174L} (*153*). On the other hand, a series of BRD4 antagonists have entered clinical trials, showing a promising prospect of alleviating MYC transcription

(*25*). This availability of drugs and therefore proven selective druggability, taken together with general advantages of dual inhibitors vs. combination therapy (cf. page 27), provided us with a promising rationale to design a dual inhibitor for ALK^{F1174L} and BRD4.

BI-2536 was selected as starting point for this study, since target specificity has already been discussed previously and it already showed some ALK affinity (Figure 18B) (*14*). Also, separated acetyl-lysine binding and hinge binding moieties promised to be tuneable in their binding properties step by step to establish a stable SAR. More specifically, the dihydropteridinone carbonyl and the methylamino group mimicked acetyl-lysine and formed the conserved hydrogen bond interaction with N140 in BRD4. In PLK1, this region interacted with the K82 side chain of the VIAK motif in PLK1 and – mediated by water molecules - the DFG backbone (Figure 17). The amine linker together with the dihydropteridinone amine formed the canonical hinge interactions in PLK1 and ZA channel interactions in BRD4. The methoxy group protruded into a PLK1-specific cavity that is available due to a smaller leucine side chain compared to larger aromatic residues in other kinases. The methylpiperidine part of the molecule faced towards the solvent channel in PLK1 and BRD4 (*14*).

The general aim was to change kinase target specificity from PLK1 to ALK^{F1174L} while maintaining the BRD4 affinity. To do so, the acetyl-lysine mimetic as well as the hinge binding motif were kept and four regions that decorated the scaffold of BI-2536 were selected for chemical modifications: R1 at the (*R*)-ethyl group, R2 at the cyclopentyl group, R3 at the methoxy group and R4 at the amide linker/methylpiperidine site (Figure 18A).

Omitting the (*R*)-ethyl group or changing it to the (*S*) enantiomer reduced affinity for PLK1, but also to similar extent for BRD4 and ALK^{F1174L}, thus (*R*)-ethyl was kept at position R1. Changing the cyclopentyl moiety in BI-2536 to 3-bromobenzyl and benzyl group maintained the BRD4 activity and led to a 30-40-fold decrease in PLK1 affinity. Although the overall ALK^{F1174L} affinity was lower than BI-2536, the large decrease in PLK1 affinity indicated that changes in this position are important means in influencing the ALK^{F1174L}/PLK selectivity. The methoxy residue at R3 was proposed to be responsible for PLK1 selectivity (*14*) and indeed, the change to an ethoxy group reduced PLK1 affinity at least 10-fold and ALK_{F1174L} only two-fold. A similar change in affinities was achieved by removing the amide linker between the hinge binding motif and the methylpiperidine. Combining the changes that led to favourable ratios in ALK^{F1174L}/PLK selectivity into one molecule led to compound B1, which featured the benzyl group at R2, the ethoxy group at R3 and the sole methylpiperidine without amide linker at R4. B1 was the first compound to exhibit superior ALK^{F1174L} affinity over PLK1 affinity.

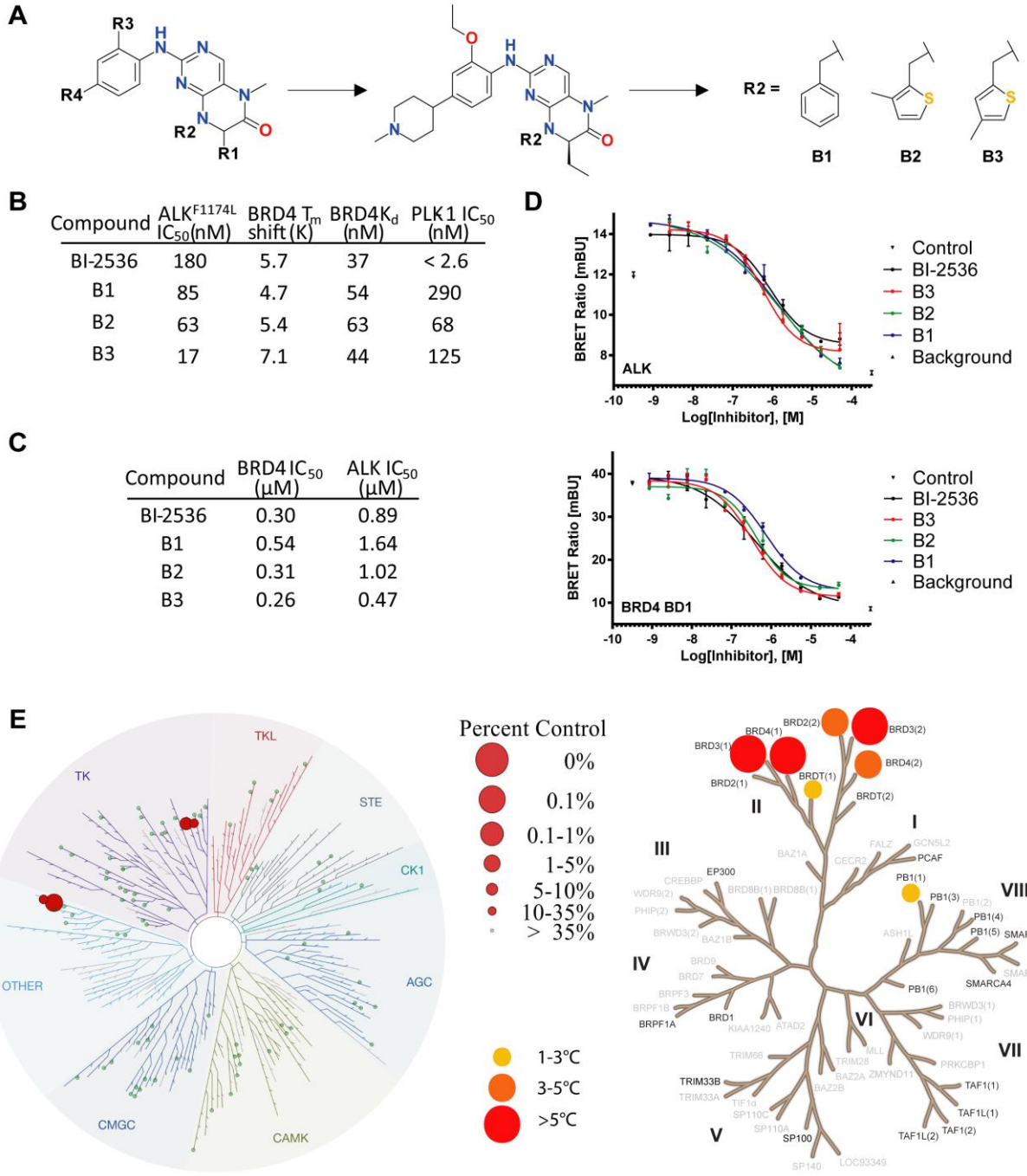


Figure 18 A The chemical scaffold of BI-2536 was used to make modifications on four positions (R1-R4). After the establishment of an SAR for positions R1-R3, three variations on positions R2 defined the lead compounds B1-B3 **B** *In vitro* data for the three lead compounds B1-B3 in comparison with BI-2536. Kinase specificity was changed from PLK1 to ALK^{F1174L}. IC$_{50}$ values for ALK^{F1174L} were obtained by LanthaScreen® Eu kinase binding assay, BRD4 thermal shift values by DSF, BRD4(1) K$_D$ values by ITC and PLK1 IC$_{50}$ values by Z´-LYTE™ activity assay (Thermo Fisher Scientific) **C** Summary of IC$_{50}$ values obtained by NanoBRET assay for ALK WT and BRD4(1) **D** NanoBRET graphs show target engagement for BI-2536 and B1-B3 by NanoBRET assay **E** Selectivity within the kinase tree is demonstrated by DiscoveRx KINOMEscan™ (scanEDGE) assay (left, upper legend) and within BRDs by thermal shift assay (right, lower legend) (Compounds in **A** and ALK$_{F1174L}$ and PLK1 IC$_{50}$ values in **B** were synthesised/obtained by E. Watts, kinase selectivity panel in **E** was adapted from (*16*))

Because the greatest change in affinities was achieved by perturbing R2, this residue was selected for further SAR studies. Keeping the (*R*)-ethyl group in combination with the acetyl-lysine mimetic maintained BRD4 activity throughout all modifications, so the main challenge remained in regulating the kinase selectivity. To rationalize further modifications on R2, co-crystallisation ALK WT with B1 was attempted. Unfortunately, the attempts were unsuccessful probably due to the still moderate affinity. Therefore, docking with Glide was used as selection method for other moieties. While different *ortho*-, *meta*- and *para*-substituents on the benzyl ring did not improve ALK^{F1174L} activity significantly, heterocycles with methyl substituents led to IC_{50} values better than 100 nM for ALK^{F1174L}, with at least 48-fold reduction in PLK1 activity compared to BI-2536. Among these, compounds B2 and B3 emerged as the most promising.

The three lead compounds B1-B3 all bind ALK^{F1174L} with higher affinity than 100 nM (IC_{50}s: B1: 85 nM, B2: 63 nM, 17 nM). That translates to a 10.6-fold affinity increase comparing BI-2536 and B3. BRD4 activity remained steady and amounted to 44-63 nM. PLK1 affinity decreases from <2.6 nM for BI-2536 to 290, 68 and 125 nM for B1-B3 respectively. Comparing BI-2536 with B3, PLK1 affinity dropped at least 48-fold. Another way to judge the success of the rational inhibitor design is the comparison of ALK^{F1174L}/PLK1 affinity ratios. For BI-2536 this value was 69.2, for B1 it was 0.3, for B2 it was 0.9 and for B3 it was 0.1. Altogether, B3 stands out as the inhibitor with best affinity towards ALK^{F1174L} and best ALK^{F1174L}/PLK1 affinity ratio (Figure 18B).

Next, we assessed BRD and kinase specificity of compound B3 (Figure 18E). In a thermal shift assay including 25 BRDs covering all families with at least one representative, only BRD3(1) and BRD3(2) were identified as considerable off-targets. Minor shifts were detected for BRDT(1) and PB1(1). In the kinase panel containing 97 kinase domains, only PLK1, insulin receptor (INSR) and PLK3 were to be mentioned as off-targets. The off-targets were consistent with our expectations, since all of them feature the same leucine residue in the hinge region that can accommodate larger moieties at R3.

In order to verify the results in a cellular context, target engagement of B1-B3 was shown in a NanoBRET experiment and in an MSD assay. In the NanoBRET experiment, all compounds displaced a labelled tracer molecule from BRD4(1) and ALK WT in HEK293T cells. Indeed, the best IC_{50} values were achieved with B3 (0.26 µM for BRD4(1), 0.47 µM for ALK WT, Figures 18C and 18D). The *in vitro* and NanoBRET results were also confirmed by the MSD assay. Phosphorylation of ALK^{F1174L} was shown in Kelly neuroblastoma cells with IC_{50} values in between BI-2536 and ceritinib (data shown in publication), which is in agreement with the discussed affinities.

We also investigated markers of mitotic arrest in G2/M phase as well as PLK1 phosphorylation at T210 and S137 in cells, which are markers of PLK1 inhibition. Western blot analysis confirmed that BI-2536 induced cell cycle arrest at 50 nM, whereas compound B3 did not cause any elevation of mitotic markers and PLK1 phosphorylations below 10 µM, which is well above concentrations that one would expect judging its cellular IC_{50} value.

The binding modes of compound B1 - as first compound in the series with better affinity for ALK^{F1174L} than for PLK1 – and B3 – as best lead compound – were examined by co-crystallisation with BRD4(1) (Figure 19). Both binding modes were in agreement with BI-2536. The dihydropteridinone carbonyl group formed a hydrogen bond with the conserved asparagine residue N140 and a water molecule that is coordinated by Y97. Another water bridge was formed between a dihydropteridinone nitrogen and Q85. The asymmetric ethyl group on R1 reached into the same hydrophobic pocket between L92, L94 and Y139 that has already been described for BI-2536 (*14*). All parts of the molecule within the binding pocket were well-accounted for by electron density.

In this project, we successfully designed a dual inhibitor for ALK^{F1174L} and BRD4 which was aimed at especially challenging cases of neuroblastoma. Guided by docking experiments with Glide, the BI-2536 chemical scaffold was changed through systematic exploration of chemical modifications to reach in vitro affinities of 17 nM for ALK^{F1174L} and 44 nM for BRD4 in compound B3. PLK1 affinity was decreased from under 2.6 nM to 125 nM. Below 10 µM, B3 did not show PLK1 inhibition-induced phenotypes like mitotic arrest in G2/M phase. Target engagement in HEK293T cells and Kelly neuroblastoma cells was demonstrated and the binding mode of compound B3 was confirmed as canonical binding mode in BRD4 by X-ray crystallography. These results made compound B3 a promising compound to learn more about the ALK^{F1174L} neuroblastoma phenotype and a possible precursor molecule for a drug.

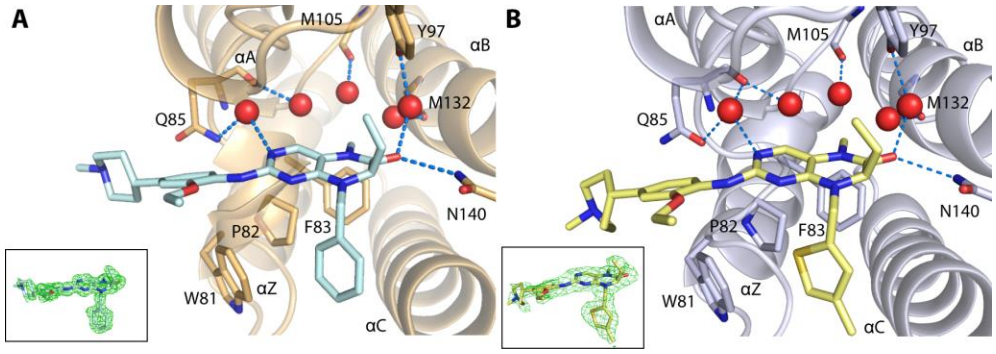


Figure 19 A Co-crystal structure of B1 with BRD4 (PDB ID: 6Q3Y) and associated associated $|2F_O|-|F_C|$ refined electron density map contoured at 1σ. b) Co-crystal structure of B3 with BRD4 (PDB ID: 6Q3Z) and associated $|2F_O|-|F_C|$ refined electron density map contoured at 1σ (Compounds B1 and B3 were synthesised by E. Watts, Figure adapted from (*16*))

Table 13 Data collection and refinement statistics of BRD4(1)-B3 and BRD4(1)-B1 complexes

Complex	**BRD4A(1)-B3**	**BRD4A(1)-B1**
PDB accession codes	**6Q3Z**	**6Q3Y**
Data Collection		
Resolution[a] (Å)	33.35-2.00 (2.05-2.00)	39.30 1.20 (1.22-1.20)
Spacegroup	*P1*	*P1*
Cell dimensions	*a* = 30.2, *b* = 39.5, *c* = 55.7 Å	*a* = 30.3, *b* = 39.5, *c* = 56.0 Å
	α = 84.4°, β = 75.2°, γ = 90.0˚	*α= 83.9°, β = 75.6°, γ =* 89.8˚
No. unique reflections[a]	15,332 (1,078)	65,153 (3,007)
Completeness[a] (%)	91.9 (88.4)	83.0 (76.9)
I/σI[a]	6.7 (3.4)	9.6 (2.7)
R_{merge}[a] (%)	0.113 (0.450)	0.055 (0.335)
Redundancy[a]	3.1 (3.0)	3.7 (3.5)
Refinement		
No. atoms in refinement (P/L/O)[b]	2,113/76/259	2,103/114/361
B factor (P/L/O)[b] ($Å^2$)	18/19/27	14/23/25
R_{fact} (%)	17.2	14.2
R_{free} (%)	23.1	17.6
rms deviation bond[c] (Å)	0.013	0.013
rms deviation angle[c] (°)	1.4	1.5
Molprobity Ramachandran		
Favour (%)	100.00	98.33
Outlier (%)	0	0
Crystallization condition	25% PEG3350, 0.2M ammonium sulphate, 0.1M bis-tris pH6.5	30% pentaerythritol ethoxylate 15/4, 0.05 ammonium sulphate, 0.1M bis-tris pH6.5

[a] Values in brackets show the statistics for the highest resolution shells.
[b] P/L/O indicate protein, ligand molecules presented in the active sites, and other (water and solvent molecules), respectively.
[c] rms indicates root-mean-square.

Targeting novel acetyl-lysine reader domains

The third aim of this thesis was the investigation of novel acetyl-lysine reader domains and their potential as drug targets. In recent years, YEATS domains emerged as newly discovered reader domains with ligand specificity covering histone acetylation as well as other histone acylation modifications such as crotonylation, butyrylation and propionylation. These acylations are presumably derived from fatty acid metabolic intermediates and are added by HATs, which are rather unselective. They represent specialised states of the cell like starvation on an epigenetic level (*38, 10*).

There are four human proteins containing YEATS domains: The two closely related ENL and AF9, YEATS2 and GAS41. All of them have been shown to recognise acylated lysines on histone tails and to be involved in a number of cancer diseases (*21, 96, 97, 100–102*). Although involvement of e.g. full length GAS41 in cancer was shown many years ago, it was recently that a direct link to the reader function of YEATS domains has been proved. Shortly after this thesis started, Wan *et. al*. and Erb *et. al*. delivered structural data on the interaction between acetylated H3K27 histone peptide and the ENL YEATS domain as well as detailed information about its involvement in progression of MLL-rearranged acute myeloid leukemia (*19, 21*).

Structural properties of the ENL YEATS binding domain

The publications treating the ENL YEATS domain reassured us in our aim to investigate YEATS domains as potential drug targets and at the same time provided us with first promising experimental approaches, so we decided to stick with ENL as our main protein of interest. This study has been published in J. Med. Chem. as Brief Article in November 2018 (*22*).

Commonly, one of the first experimental steps in our lab is the screening of inhibitor libraries with DSF. The examination of a diverse set of compounds provides us with a quick overview over the type of scaffolds that meet the binding pocket requirements. In the beginning of the project, we tried to investigate a number of established BRD4 compounds, lysine derivatives and a set of commercially available compounds that were selected with a virtual screening method similar to that which was used to identify A1 as a hit for PB1(5) (*13*). The latter was a logical choice since we dealt with so far untargeted domains and might have had to develop completely new chemical scaffolds. Unfortunately, the well-established DSF workflow proved to be impractical due to a combination of three reasons: First, the melting temperatures of the respective domains were comparably high (Table 14). Second, YEATS domains are relatively small and ligand binding is not accompanied by a large conformational change as

it is the case for e.g. kinase domains. Larger conformational changes often correlate with higher thermal shifts. Third, hit compounds are usually in the low micromolar range and need to be applied at concentrations between 100-500 µM, where intrinsic compound fluorescence started to interfere with SYPRO orange fluorescence.

Table 14 Melting points of the four human YEATS domains

Protein	ENL	YEATS2	AF9	GAS41
Melting point (°C)	62	58	52	58

Therefore, instead of selecting compounds by virtual screening and depending on their biophysical characterisation, we focused on a structure-based approach. This approach depends on establishing a stable crystallisation system that is suitable for soaking compounds. Similar strategies have been pursued successfully in fragment studies for the BRDs of ATAD2 and PCAF (*154, 155*).

In order to obtain more information about general properties of the ENL YEATS domain, it was crystallised in the ligand-free state (apo) (Figure 20, PDB ID: 6HQ0). All loops as well as side chains in vicinity of the binding pocket were accounted for by electron density. Two conserved water molecules were visible in the binding pocket. The first water molecule was coordinated by the backbone carbonyl of A79 and visible in the peptide-bound structure (PDB ID: 5J9S) (*19*). The second water molecule was nearer to the aromatic triad and coordinated by the backbone amine of Y78. Surprisingly, two alternative conformations of side chain Y78 were observed, suggesting flexibility of this side chain in absence of a ligand. The first conformation (from now on "'in' conformation") resembled so far published conformations and faces towards loop 4, forming a "bridge" over the elongated binding pocket. The second conformation (from now on "'out' conformation") was stabilised by a hydrogen bond between the tyrosine hydroxyl group and the backbone carbonyl group of E26 in loop 1. All other residues, especially F28 and F59 of the aromatic triad were clearly resolved in one conformation. The apo crystals were reproduced in the same space group. The solvent channel in the crystal structure seemed relatively large and flexibility of Y78 suggested suitability for soaking experiments.

To gain first insights into binding properties of other ligands and to examine the discovered flexibility of Y78, soaking experiments with acetyl-lysine into apo crystals were attempted. Interestingly, the data set yielded electron density at the position where the second water molecule (at Y78) is positioned, but its shape was larger and extended towards the entrance of the pocket and the S58 side chain (Figure 21A, PDB ID: 6HPZ). Indeed, the density fit the head of acetyl-lysine, which formed two hydrogen bonds towards the backbone nitrogen atoms of Y78 and A79.

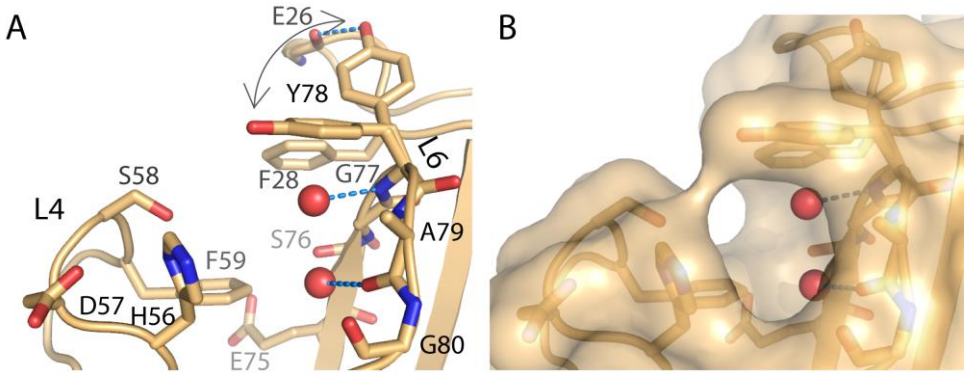


Figure 20 Structure of apo ENL YEATS domain (PDB ID: 6HQ0) **A** Binding site of apo ENL YEATS domain with Y78, a residue part of the key interacting aromatic triad, is flexible between "in" and "out" conformations. Bound water molecules within the pocket are shown in red spheres **B** Surface representation of the binding site of ENL (derived from (*22*))

Table 15 Data collection and refinement statistics for apo ENL YEATS domain

Complex	**apo ENL YEATS**
PDB accession code	**6HQ0**
Data Collection	
Resolution (Å)	46.12-1.81 (1.84-1.81)
Spacegroup	$P4_32_12$
Cell dimensions	a=b= 49.4, c = 128.3 Å
	α=β=γ=90.0˚
No. unique reflections	15,436 (895)
Completeness (%)	100.0 (99.8)
I/σI	14.3 (3.2)
R_{merge} (%)	0.090 (0.841)
Redundancy	11.9 (10.2)
Refinement	
No. atoms in refinement (P/L/O)	1,189/0/93
B factor (P/L/O) (Å^2)	35/0/45
R_{fact} (%)	19.2
R_{free} (%)	23.7
rms deviation bond (Å)	0.015
rms deviation angle (°)	1.5
Molprobity Ramachandran	
Favour (%)	97.16
Outlier (%)	0
Crystallization condition	30% 2000MME, 0.1M potassium bromide

The rest of the compound towards Cα is usually integrated into the histone protein and fixed in place. In this case, the missing electron density even after refinement strongly suggested averaging and high flexibility. Interestingly, binding of acetyl-lysine did not cause the Y78 side chain to adopt an in conformation. Both in and out conformations were present.

An overlay of the acetyl-lysine structure with the shown apo-structure and the published peptide-bound structure revealed further details about the behaviour of the binding pocket residues (Figure 21B).

Several elements attracted attention in the superimposed structures. The first element that caught the eye is the water molecules. The conserved water molecule coordinated by the carbonyl group of A79 was at the same position in all three structures. The second water molecule in the apo-structure (in blue) however

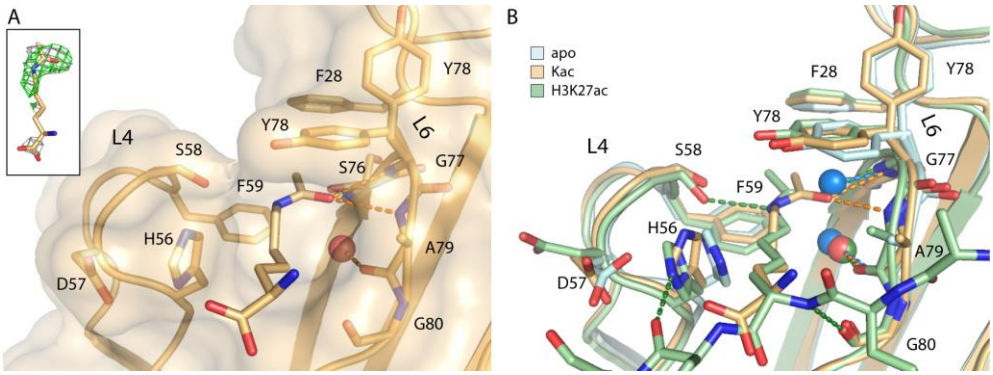


Figure 21 Structure of ENL in complex with acetyl-lysine **A** Interaction between the bound acetyl lysine within ENL YEATS (PDB ID: 6HPZ). Inset shows $|F_O|$-$|F_C|$ omitted electron density map contoured at 3σ (green) and $|2F_O|$-$|F_C|$ refined map contoured at 1σ (grey) for acetyl-lysine **B** Overlay of apo, acetyl-lysine-bound, and H3K27ac-complexed (PDB ID: 5J9S) (*19*) structures revealed a conserved binding mode of acetyl-lysine that displaced a water for a hydrogen bond interacting with loop 6 backbone amines (adapted from (*22*))

Table 16 Data collection and refinement statistics for ENL YEATS domain in complex with acetyl-lysine

Complex	**ENL with acetyl-lysine**
PDB accession code	**6HPZ**
Data Collection	
Resolution (Å)	64.80-2.30 (2.38-2.30)
Spacegroup	$P4_32_12$
Cell dimensions	a=b= 49.3, c = 129.6 Å
	α=β=γ=90.0˚
No. unique reflections	7,707 (730)
Completeness (%)	99.9 (100.0)
I/σI	9.9 (2.1)
R_{merge} (%)	0.147 (0.992)
Redundancy	8.6 (8.6)
Refinement	
No. atoms in refinement (P/L/O)	1,185/13/41
B factor (P/L/O) (Å^2)	41/73/40
R_{fact} (%)	23.3
R_{free} (%)	29.6
rms deviation bond (Å)	0.009
rms deviation angle (°)	1.3
Molprobity Ramachandran	
Favour (%)	97.87
Outlier (%)	0
Crystallization condition	25% PEG 1500

coincided with the position of the acetyl carbonyl groups in the acetyl-lysine-bound structure (grey, Kac) and the peptide structure (green), indicating it as an important characteristic interaction in ENL YEATS. This importance was further shown by the pronounced electron density at this position in the acetyl-lysine structure. Second, the side chains underwent a concerted movement upon ligand binding. Especially the side chains of H56, D57 and S58 moved more inward the larger the ligand gets. The S58 side chains acted as beta-sheet mimetic counterpart of the aforementioned loop 6 backbone interaction, forming a hydrogen bond with the acetyl-lysine nitrogen of the peptide-structure. Third, although slight movement of residues F28 and F59 of the aromatic triad for accommodation of the ligand was visible,

substantial movement of Y78 was evident in the superimposition. Most interestingly, despite their highlighted role as inducers of the π-π-π stacking interaction, sole binding of acetyl-lysine did not fix this side chain in the "in" conformation. Using the knowledge from the two novel structures, a selection strategy for systematic fragment soaking was devised.

ENL YEATS fragment study

Being intrigued by the unique behaviour of the binding pocket in the apo/acetyl-lysine-bound state, we decided on possible properties for ligands that bind to the ENL YEATS pocket. Hydrogen bonds with the backbone of loop 6, either by water molecules or the acetyl carbonyl group, dominated the published as well as the newly obtained structures. Also, the S58 side chain was in suitable position to engage with the acetyl nitrogen in a hydrogen bond. However, binding of acetyl-lysine itself was not sufficient to induce the Y78 side chain into "in" position. The working hypothesis was that the formation of hydrogen bonds is most essential and the often discussed π-π-π interaction with the aromatic triad served as a means to increase ligand affinity.

Therefore, our strategy was to use compounds with an amide bond as core element for soaking. Comparison of the H3K27ac-complexed ENL structure (PDB ID: 5J9S) with a H3K9cr-complexed AF9 structure (PDB ID: 5HJB) showed that the rear of the binding pocket offers space to decorate the amide bond with an elongated residue, potentially interacting with the π-electron systems of F59, F28 and Y78 (Figure 22A). We termed the residue attached to the carbonyl side of the amine R1. We also considered an R2 residue on the amine side of the amide bond, as the fragment is not attached to a peptide backbone that determines its orientation. Thus, the amide might be flipped (e.g. the carbonyl can potentially interact with backbones of G77, Y78 or A79), in which case R2 would face to the rear pocket (Figure 22B).

We selected a total of 19 compounds that were either commercially available (C1-C10, C13-C19) or synthesised by collaborators Jurema Schmidt and Daniel Merk (C11 and C12) and grouped them into five groups (Figures 23A and Supplementary Figure S4). Denotation of R1 and R2 refer to Figure 22. All compounds of this fragment study contained an amide with one or two flanking aromatic residues. Group I contained an aromatic group on R2 (C1-C7), group II contained an aromatic moiety on R1 (C8), group III contained aromatic moieties on R1 and R2 (C9-C14), group IV contained an aromatic group on R2 and an aromatic moiety on R1 connected by a spacer atom and group V contained an aromatic group on R1 and an aromatic substitution on R2 connected by a spacer atom.

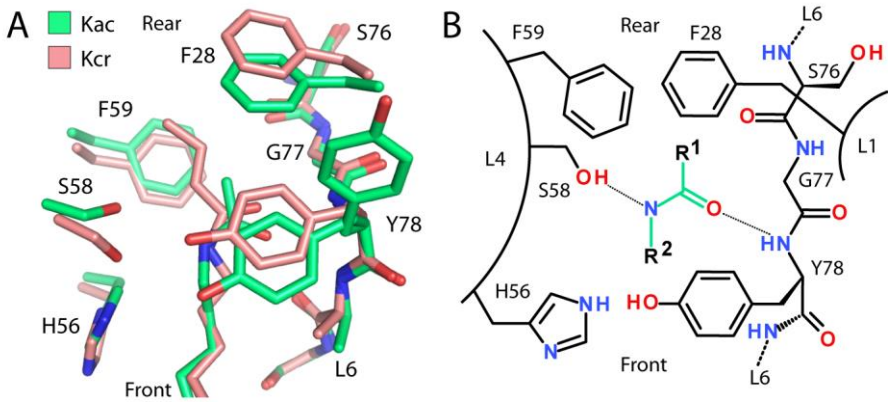


Figure 22 Analysis of the ENL YEATS binding pocket **A** Superimposition of ENL-H3K27ac (Kac) structure (PDB ID: 5J9S) with AF9-H3K9cr (Kcr) structure (PDB ID: 5HJB) shows space in the rear pocket that can be used to involve aromatic triad in inhibitor interaction (*19, 17*) **B** Scheme depicting ENL binding pocket with amide bond as fragment core decorated by groups R1 and R2 (adapted from (*22*))

Compounds C1-C19 were examined with DSF, ITC and used for soaking into apo ENL YEATS crystals. The DSF measurements did not yield useful results for reasons discussed above (cf. Table 14). The ITC measurements did not lead to sigmoidal plots, presumably due to weak interaction, but binding heat was detectable for some compounds (Figure 23 and Supplementary Figure S4). Unfortunately, some compounds diminished diffraction quality of the ENL crystals, but datasets were obtained for crystals soaked with compounds C1, C2, C3, C9, C10, C11, C12, C14, C15, and C19.

Analysis of unrefined datasets showed extended densities in close proximity to the Y78 backbone, where the acetyl carbonyl group forms a hydrogen bond with the backbone amine. Most of the densities resembled the density of acetyl-lysine, strongly suggesting the presence of the ligands. However, most of the densities did not stretch to the front and the rear of the binding pocket, making modelling of the ligands impossible (Supplementary Figure S5). Nevertheless, in cases C12 and C19, the densities could be modelled and refined and allowed deciphering the binding modes (Figure 23B).

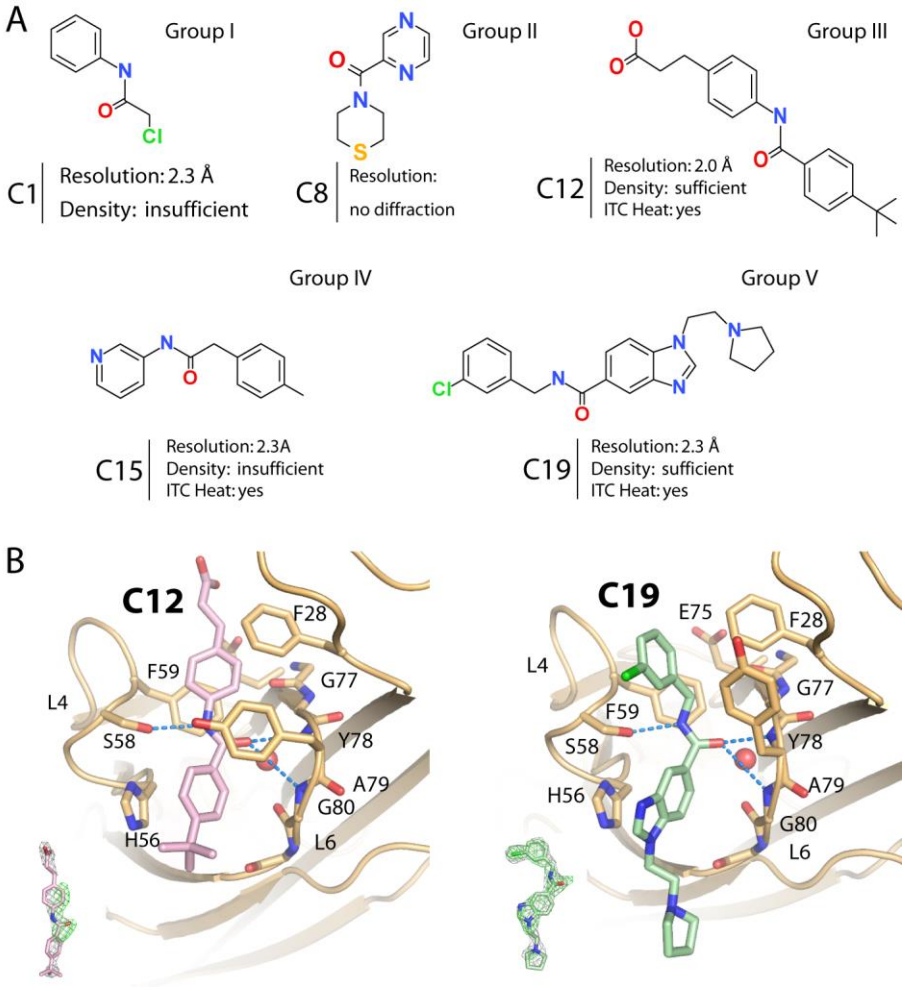


Figure 23 Crystallography-based fragment study on ENL YEATS domain **A** Examples of selected fragments for groups I-V **B** Structures of ENL in complex with 12 (left, PDB ID: 6HPY) and 19 (right, PDB ID: 6HPX). Insets show $|F_O| - |F_C|$ omitted and $|2F_O| - |F_C|$ electron density map contoured at 3σ (green) and 1σ (gray), respectively. Hydrogen bonds are shown as dashed lines and water molecules as red spheres (Compound C12 was synthesised by J. Schmidt, Figure adapted from (*22*))

The amide bond of compound C12 adopted an inverted orientation compared to the ligands containing acetyl lysine. Hydrogen bonds were formed between the S58 side chain and amide nitrogen as well as between amide carbonyl and the backbone amides of residues Y78 and A79. The aromatic moiety on R1 adopted a parallel conformation to H56, forming a π-π-stacking interaction (distance between 3.4 and 4.4 Å). The R2 aromatic group was positioned in the ENL aromatic triad and brings Y78 into the "in" conformation. The carboxyl moiety did not specifically interact with the protein (nearest residue is E75 with 4.2 Å). The amide bond of C19 adopted an identical orientation and interacted in a similar way with H56. However, the chlorine-substituted aromatic group at R2 was flipped and faced outward of the pocket, forcing Y78 into an "out" conformation.

Table 17 Data collection and refinement statistics of ENL YEATS-C12 and ENL YEATS-C19 complexes

Complex	**ENL-C12**	**ENL-C19**
PDB accession codes	**6HPY**	**6HPX**
Data Collection		
Resolution[a] (Å)	66.36-2.00 (2.05-2.00)	45.80-2.30 (2.38-2.30)
Spacegroup	$P4_32_12$	$P4_32_12$
Cell dimensions	a=b=49.0, c=132.7 Å	a=b=48.4, c=137.4 Å
	α=β=γ=90.0˚	α=β=γ=90.0˚
No. unique reflections[a]	11,691 (850)	7,846 (719)
Completeness[a] (%)	100.00 (100.00)	100.00 (100.00)
I/σI[a]	15.7 (7.7)	15.3 (7.0)
R_{merge}[a] (%)	0.079 (0.216)	0.057 (0.169)
Redundancy[a]	8.6 (8.9)	5.6 (5.8)
Refinement		
No. atoms in refinement (P/L/O)	1,179/24/79	1,184/27/45
B factor (P/L/O) (Å^2)	39/64/45	41/63/44
R_{fact} (%)	21.7	21.1
R_{free} (%)	26.0	26.1
rms deviation bond (Å)	0.013	0.013
rms deviation angle (°)	1.4	1.3
Molprobity Ramachandran		
Favour (%)	93.62	94.37
Outlier (%)	0	0
Crystallization condition	25% PEG3350, 0.2M ammonium sulphate, 0.1M bis-tris pH5.5	25% PEG Smear Medium, 0.1M citrate pH5.5

[a] Values in brackets show the statistics for the highest resolution shells.

First, the C12- and C19-complexed structures proved the hypothesis that a core amide bond was suitable to mimic the natural acetyl-lysine interaction. Second, by showing both conformations of Y78, they confirmed that the stacking mechanism may be preferential, but is not obligatory for ENL YEATS inhibitors. Third, they identified a π-stacking interaction with H56 as beneficial interaction. ITC measurements of both compounds C12 and C19 yielded binding heat, but the data did not allow a reliable fit, so only an estimation of an affinity range between 20-50 µM was made (Figure 24B).

Two questions drove subsequent research in the area of novel ENL YEATS inhibitors: 1) What factors increase the affinity of low-micromolar hits like C12 and 19 to (sub-)micromolar? 2) Are there alternative moieties to the amide core which can serve as acetyl-lysine mimetics? Question 1 was

partially answered by the rest of this fragment study and partially by the discovery of an ENL/AF9 YEATS chemical probe, that was a collaborative effort between SGC Oxford and SGC Frankfurt (*22, 24*). Question 2 was addressed by so far unpublished soaking experiments with alternative scaffolds, being described on page 100 (Figure 27).

The question of more affine inhibitors lead us to a subset of benzimidazole derivatives that were part of a larger screening study targeting YEATS domains (*23*). The compounds featured an amide core and flanking aromatic residues, falling within group III (cf. Figure 23A). ITC experiments revealed affinities ranging from 0.8 µM to 2.4 µM (Figure 24A and Supplementary Figure S6). The higher the amount of electrons at the R1 substituent, the higher the affinity tended to be. Being curious about the 10-20-fold affinity increase compared to C12, we solved the crystal structure of the complex between ENL YEATS with C20, which had the highest affinity (807 nM) (Figure 24B). C20 adopted the same orientation as C19 and C20 with R1 facing the front and R2 facing the rear of the pocket. The amide group formed the known hydrogen bonds with the backbone of Y78/A79 and S58. The R1 aromatic group with iodine substitution adopted a planar orientation towards H56. The difference to compounds C12 and C19 is evident when looking at the rear pocket. The benzimidazole group and its piperidine extension formed additional hydrogen bonds. The benzimidazole amine interacted with the backbone carbonyl group of S76 and the piperidine extension bent slightly downwards to interact with the E75 side chain. This hydrogen bond pattern reduced the degrees of freedom and offers optimal π-electron orientation that rotated the Y78 side chain into the “in” conformation. The differences in affinity within the subset could be explained by different degrees of electron density in the aromatic moiety that participated in the π-stacking interaction with H56. The increase of hydrogen bonds was well-represented in the gain of ITC heat rates. Approximately three times as much heat originated from the first injections comparing compound C20 with C12/C19 (Figure 24B).

Our structural studies revealed so far undiscovered features of ENL YEATS. Two water molecules were conserved in the cave-like binding pocket, one of which is depleted and mimicked by acetyl-lysine histone peptides. The apo-structure exposed an intrinsic flexibility of Y78 with “in” or “out” conformation. Fragment-complexed structures with C12 and C19 showed a flipped amide forming hydrogen bonds to the loop 6 backbone and the S58 side chain. The R1 aromatic substitutions adopted a planar π-stacking conformation with H56. Considerable differences were evident in the interactions of the rear part of the pocket. C12 adopted a planar stacked with a Y78 “in” conformation, whereas C19 faced upwards and forced Y78 “out”. Means to increasing the affinity was shown with the crystal

structures of ENL YEATS with C20, which formed additional hydrogen bonds with its benzimidazole amine and the piperidine amine.

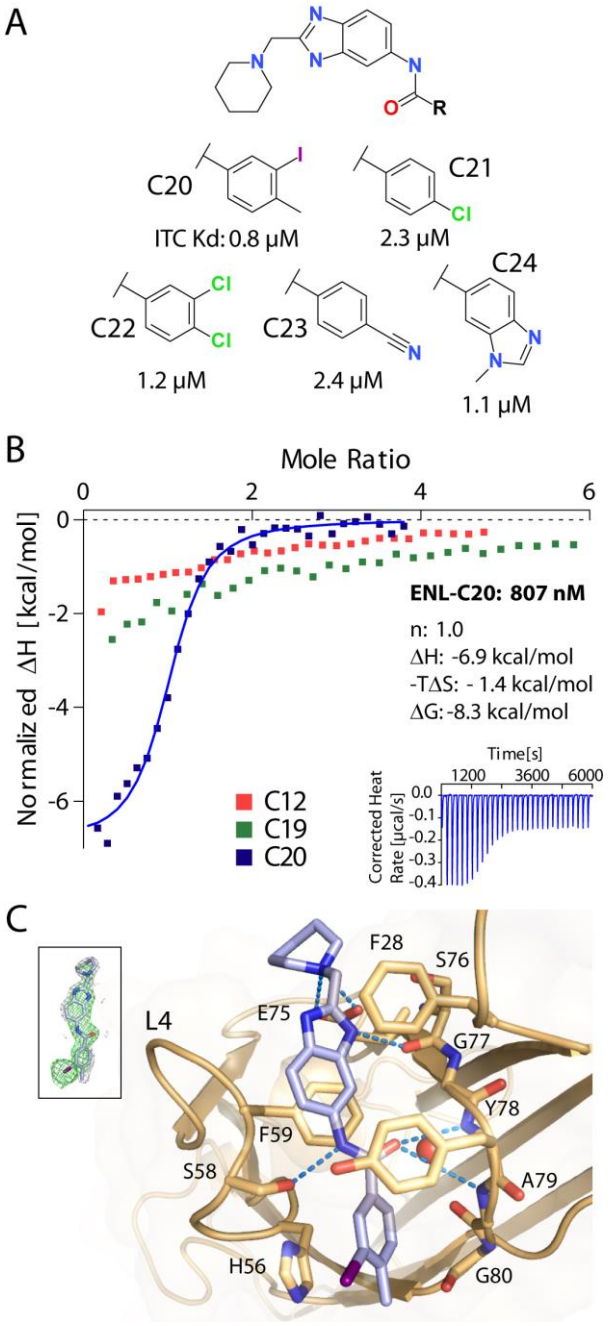


Table 18 Data collection and refinement statistics of ENL YEATS-C20 complex

Complex	**ENL with C20**
PDB accession code	**6HPW**
Data Collection	
Resolution (Å)	48.74-1.90 (1.94-1.90)
Spacegroup	$P4_32_12$
Cell dimensions	a=b=48.7, c=131.6 Å α=β=γ=90.0˚
No. unique reflections	13,292 (846)
Completeness (%)	100.0 (100.0)
I/σI	10.8 (2.5)
R_{merge} (%)	0.115 (0.855)
Redundancy	12.8 (12.8)
Refinement	
No. atoms in refinement (P/L/O)	1,180/27/91
B factor (P/L/O) (Å^2)	45/42/49
R_{fact} (%)	21.3
R_{free} (%)	28.1
rms deviation bond (Å)	0.014
rms deviation angle (°)	1.5
Molprobity Ramachandran	
Favour (%)	97.16
Outlier (%)	0
Crystallization condition	25% PEG 3350, 0.2M ammonium acetate, 0.1M bis-tris pH5.5

Figure 24 Benzimidazole derivatives and characterisation of the interaction of C20 with ENL **A** Chemical structures of a set of benzimidazoleamide derivatives and their ITC K_D values. **B** Crystal structure of the ENL YEATS in complex with C20 (PDB ID: 6HPW). Insets show $|F_O|-|F_C|$ omitted electron density map contoured at 3σ (green) and $|2F_O|-|F_C|$ refined map contoured at 1σ (grey). **C** ITC normalised binding heat of the interactions between C12, C19, and C20 with ENL. Inset shows the isotherm of raw injection heats for C20 (Compounds C20-C24 were synthesised by M. Moustakim, Figure derived from (*22*))

Discovery of an ENL/AF9 YEATS Domain Chemical Probe

The benzimidazole series were improved in collaboration with SGC Oxford and led to the chemical probe SGC-iMLLT (Figure 25). This work was published in Angewandte Chemie in November 2018 (*24*).

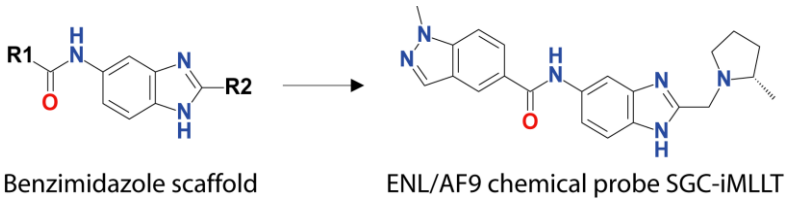


Figure 25 Benzimidazole scaffold and its development to the ENL/AF9 chemical probe SGC-iMLLT (Synthesis by M. Moustakim)

The optimisation of the benzimidazole scaffold by our colleagues at the University of Oxford focused on attachments on the carbonyl side of the amine (R1) and modifications on the C2 atom of benzimidazole. The modifications led to SGC-iMLLT, which bound to ENL with a K_D of 129 nM and to AF9 with a K_D of 77 nM (determined by ITC). The values were in good agreement with the IC_{50} of 270 nM determined by NanoBRET (measured by colleagues). The binding mode generally resembled the binding mode of C20 (Figure 26A, cf. Figure 24C) including the orientation of the central amide, the stacking of R1 and the hydrogen bonds formed by the benzimidazole moiety and the (*S*)-2-methyl-pyrrolidine (with the E57 side chain). The crystal structures in Figure 26 were solved by me. The differences were limited to a larger heterocycle on R1 as well as the 2-methyl-pyrrolidine that bound to the back pocket. However, the co-crystal structure did not provide a trivial explanation for the increase in affinity. The *R* enantiomer of SGC-iMLLT was also synthesised, but it was approx. eight times less active in AlphaScreen and ITC experiments. An investigation about how fast SGC-iMLLT is metabolised in primary human hepatocytes revealed a half-life $t_{1/2}$ of 53 min (half-life measurements done by colleagues). In order to make SGC-iMLLT less prone to N-demethylation, compound C94 was synthesised. Its binding mode essentially corresponded with the binding mode of SGC-iMLLT except that the heterocycle on R1 adopted a 180° flipped conformation to accommodate the newly introduced cyclopropyl group. Nevertheless, despite the retained affinity (ENL K_D 58 nM), C94 showed shorter half-life of approx. 30 min.

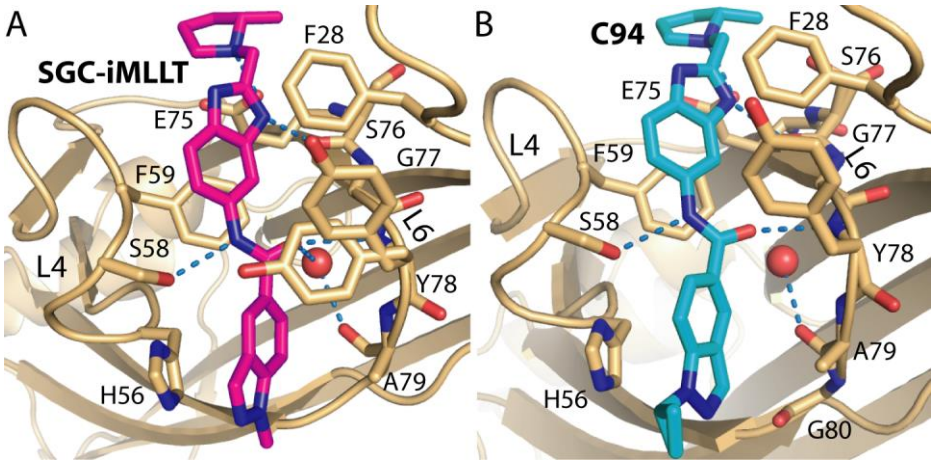


Figure 26 SGC-iMLLT and its derivative C94 in complex with ENL YEATS **A** SGC-iMLLT in complex with ENL YEATS (PDB ID: 6HT1). Y78 adopted two different conformations. Despite the core amide interaction, the methyl-benzimidazole at R1 stacked with H56 and the benzimidazole at R2 formed a hydrogen bond and interacted with the aromatic triad. The 3-(*S*)-methylpyrrolidine extension enhanced activity by interacting with the E75 side chain **B** Compound C94 (PDB ID: 6HT0) was synthesised to improve metabolism properties and bound highly similar to SGC-iMLLT (Compounds were synthesised by M. Moustakim, data published in (*24*))

Target engagement was investigated by colleagues with a variety of methods (*24*). Endogenous ENL from MV4;11 cells that were incubated with SGC-iMLLT showed higher thermal stability than in case of incubation with its inactive control (*R* enantiomer). The inactive control did not show effects up to 10 µM incubation concentration. Fluorescence recovery was measured with GFP-tagged WT ENL or AF9 in a FRAP experiment. The half recovery times $t_{1/2}$ were relatively fast with 0.46 s and 0.6 s respectively, which meant that the majority of protein was anyway mobile and the effect of an inhibitor disrupting the histone interaction would not have made a noticeable difference. Therefore, the cells were pre-incubated with 2.5 µM SAHA to increase global histone acetylation. This increased the initial $t_{1/2}$ for ENL to 0.78 s and for MLLT3 to 1.32 s. Incubation with 1 µM SGC-iMLLT led to faster recovery times of 0.46 s and 0.56 s respectively, which indicated that the YEATS-histone interaction was disrupted. Correspondingly, the more inactive control *R* enantiomer did not lead to faster recovery times. Surprisingly, treatment of HEK cells with SGC-iMLLT failed to displace the tracer molecule from a full-length ENL:NanoLuc protein in a NanoBRET experiment. On the other hand, displacement in a concentration-dependent manner could be shown in case of AF9:NanoLuc. Lastly, a gene profiling verified the potential of SGC-iMLLT as a drug development tool in the context of AML tumorigenesis. SGC-iMLLT diminished expression of MAY and DDN and increased the expression of CD86 in comparison to DMSO.

Table 19 Data collection and refinement statistics of ENL YEATS-SGC-iMLLT and ENL YEATS-C94 complexes

Complex	**ENL-SGC-iMLLT**	**ENL-C94**
PDB accession codes	**6HT1**	**6HT0**
Data Collection		
Resolution[a] (Å)	48.84-2.10 (2.17-2.10)	45.77-1.80 (1.84-1.80)
Spacegroup	$P4_32_12$	$P4_32_12$
Cell dimensions	a=b=48.8, c=133.2 Å	a=b=48.8, c=132.9 Å
	α=β=γ=90.0˚	α=β=γ=90.0˚
No. unique reflections[a]	10,098 (956)	15,693 (900)
Completeness[a] (%)	100.00 (100.00)	100.00 (100.00)
I/σI[a]	13.2 (3.4)	17.6 (2.6)
R_{merge}[a] (%)	0.069 (0.463)	0.059 (0.774)
Redundancy[a]	7.8 (7.6)	10.0 (10.3)
Refinement		
No. atoms in refinement (P/L/O)	1,200/29/89	1,221/31/86
B factor (P/L/O) (Å^2)	53/59/61	35/41/44
R_{fact} (%)	20.1	20.1
R_{free} (%)	25.2	24.9
rms deviation bond (Å)	0.015	0.016
rms deviation angle (°)	1.5	1.6
Molprobity Ramachandran		
Favour (%)	96.45	97.86
Outlier (%)	0	0
Crystallization condition	25% PEG3350, 0.2M ammonium sulphate, 0.1M bis-tris pH5.5	25% PEG Smear Medium

[a] Values in brackets show the statistics for the highest resolution shells.

Alternative chemical scaffold for YEATS domains

One aim in the research of chemical probes is the development of at least two structurally diverse chemical scaffolds. Diverse scaffolds targeting a binding pocket with similar selectivity help to understand which phenotypical aspects are related to off-target effects. While the two previous sections characterised the binding pocket as flexible and druggable and proved that nanomolar selective inhibitors can be developed for this new type of domain, this last section deals with alternative chemical scaffolds.

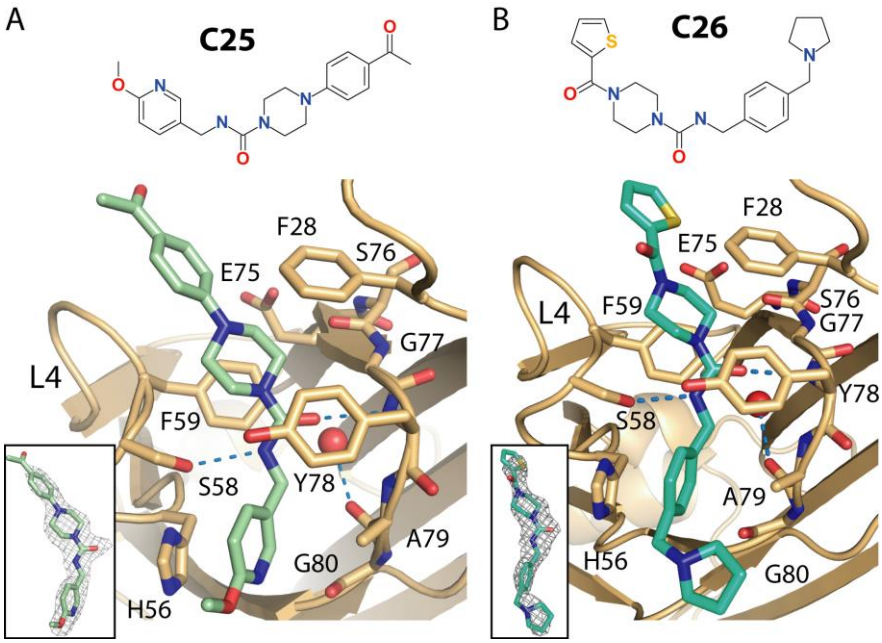


Figure 27 YEATS urea inhibitors C25 and C26 in complex with ENL YEATS **A** C25 formed amide bond-like interactions with the S58 side chain and the L6 backbone. The saturated piperidine analogue faced towards the rear pocket where usually aromatic decorations interact with the aromatic triad. The methoxypyridine occupied the front pocket and adopted a planar conformation to H56. The Y78 side chain was in an "in" conformation. $|2F_O| - |F_C|$ electron density map contoured at 1σ shown in inlet **B** Compound C26 essentially formed the same interactions with the saturated heterocycle accommodated in the rear pocket. $|2F_O| - |F_C|$ electron density map contoured at 1σ shown in inlet (data unpublished)

Screening efforts that identified the benzimidazole compounds as ENL/AF9 inhibitors (*23*) also led to the discovery of compounds C25 and C26. As opposed to SGC-iMLLT, they were characterised by a central urea moiety, flanked by one aromatic group and one saturated heterocycle. The affinity of C25/C26 to ENL YEATS was estimated to be in the low micromolar range by SPR measurements (performed by colleagues, data not shown). Since these compounds represented a different scaffold and had moderate affinity, we solved the crystal structures with ENL YEATS to learn more about the binding features (Figure 27). Both compounds essentially showed the same binding characteristics. The core urea moiety acted as acetyl-lysine mimetic and interacted with the S58 side chain and the backbone of loop 6 in a similar manner as the amide bond of SGC-iMLLT. Towards the front pocket, the flanking aromatic substitutions adopted a planar conformation to the H56, which we observed so far for all inhibitory fragments and affine inhibitors. Interestingly, the saturated heterocycles, that flanked the urea moiety from the other side, were oriented towards the rear pocket, where the π-π-π-stacking interactions with the aromatic triad F28-F59-Y78 have been reported (*19*).

Table 20 Data collection and refinement statistics of ENL YEATS-SGC-iMLLT and ENL YEATS-C94 complexes

Complex	**ENL-C25**	**ENL-C26**
Data Collection		
Resolution[a] (Å)	48.94-1.80 (1.85-1.80)	46.02-1.95 (2.00-1.95)
Spacegroup	$P4_32_12$	$P4_32_12$
Cell dimensions	a=b=48.9, c=132.8 Å	a=b=49.2, c=131.0 Å
	α=β=γ=90.0˚	α=β=γ=90.0˚
No. unique reflections[a]	15,750 (900)	12,442 (831)
Completeness[a] (%)	100.00 (100.00)	99.9 (99.1)
I/σI[a]	12.9 (1.9)	8.4 (1.4)
R_{merge}[a] (%)	0.072 (0.790)	0.093 (0.726)
Redundancy[a]	8.6 (8.9)	7.0 (5.9)
Refinement		
No. atoms in refinement (P/L/O)	1,172/27/72	1,191/29/61
B factor (P/L/O) (Å^2)	43/50/49	29/79/57
R_{fact} (%)	22.4	21.8
R_{free} (%)	25.4	28.2
rms deviation bond (Å)	0.016	0.014
rms deviation angle (°)	1.3	1.4
Molprobity Ramachandran		
Favour (%)	97.12	98.58
Outlier (%)	0	0
Crystallization condition	25% PEG3350, 0.2M sodium chloride, 0.1M bis-tris pH6.5	25% PEG3350, 0.2M sodium chloride, 0.1M bis-tris pH6.5

[a] Values in brackets show the statistics for the highest resolution shells.

Also, the Y78 side chain adopted an “in” conformation both in the C25 structure (Figure 27A) and the C26 structure (Figure 27B) despite of the lack of aromatic interactions in the rear pocket and the higher steric requirements of the space-consuming ring conformation. We refer to this compound conformation as a “flipped orientation”, since we expected the flanking aromatic moieties located in the front pocket to be positioned in proximity to the aromatic triad. This flipped orientation has not been reported so far and is therefore a particular feature of these urea scaffolds with an aromatic/non-aromatic-flanking substitution. The orientation was clearly assigned by electron density for both compounds and the resulting binding mode was associated with minimum clashes with binding pocket side chains (Figure 27 inlets).

Despite the similarity between the urea moiety in C25 and C26 and the amide bond in SGC-iMLLT, the data demonstrated the possibility to develop YEATS inhibitors using alternative chemical scaffolds which one the one hand enabled other types of modifications and on the other hand enriched the understanding of the binding pocket properties.

Discussion

Inhibitors for untargeted bromodomains

In the article "Discovery and Optimization of a Selective Ligand for the Switch/Sucrose Nonfermenting-Related Bromodomains of Polybromo Protein-1 by the Use of Virtual Screening and Hydration Analysis", we described the discovery of inhibitors for the BRD family VIII (*13*).

The approach consisted of a unique combination of computational methods, biophysical profiling and chemical synthesis. The study illustrates well how this methodological combination reduces the synthesis efforts for improving the affinity of a ligand by selection of suitable starting points by considering a vast variety of small molecules from *in silico* libraries. It is therefore an approach well-suitable for laboratories without access to high-throughput screening facilities, which are costly due to the purchase of compound libraries and devices/consumables used for the screenings. Similar hybrid virtual screening approaches have been established as web servers and are available online (*156*). However, since known BRD ligands were used to assess the similarity analysis of *in silico* library molecules in the first virtual screening step, there is a certain bias that might have hampered the discovery of new chemical scaffolds to some degree. So, applying the method to cases where no inhibitory fragments are available, as it was the case for the YEATS domains at the beginning of this PhD in 2016, is likely to yield many false positive results. Nevertheless, the minimalistic synthetic effort led to a ~3-fold affinity improvement from compound A1 to compound A3, and contributed to the development of PFI-3 which reached 48 nM affinity towards PB1(5) (*89*). Overall, the approach is suitable for cases where first inhibitors are available and new rationales for improvement of inhibitors are being sought after. In this study, the discovered unstructured water molecules in the periphery of the ZA channel was a characteristic for family VIII, hence the result were selective inhibitors for this family. The hydration of binding pockets is generally accepted as an important characteristic to be considered in ligand design (*157*). Despite that the calculation of hydration maps is under constant development and generally considered as fairly accurate, the prediction of affinities for inhibitors which substitute water molecules are still inaccurate and remains challenging (*158*).

Three BRD families are worth highlighting because of their physiological importance and because inhibitors are so far underrepresented (cf. Figure 5). Family V contains nuclear autoantigen SP100 and nuclear body proteins SP110 and SP140 (*8*). All the respective bromodomains contain an atypical

tyrosine at a position where asparagine is conserved in BRDs (N140 in BRD4). SP100 is a transcriptional coactivator of cytokine gene regulator Protein C-ets-1 (ETS1) through which it might play a role in angiogenesis (*159*). It is also a constituent of promyelocytic leukemia protein (PML) nuclear bodies which are involved in processes like cell growth, differentiation and apoptosis (*160*). The mechanism through which PML bodies and the SP100 tandem PHD/BRD in particular exert their function remains unclear. Research aiming at the function of the bromodomain would therefore benefit from the development of a chemical probe. The tyrosine mutation occupies a large fraction of the usual biding site, making the SP100 pocket flatter than most BRDs. Beside the BAZ proteins, Family VI consists mainly of TRIM proteins, which are characterised by a RING finger domain, one or two B-box motifs and a coiled-coil region (*161*). For instance, TRIM24 comprises an N-terminal trim domain, a nuclear receptor interaction motif and a C-terminal PHD/BRD. They regulate cellular processes like apoptosis, cell cycle and viral response, hence aberrations entail diseased conditions of cells (*8*)(*156*). Although the druggability of TRIM24, TRIM33B and TRIM66 was rendered as difficult, their open ZA loop configuration is similar to the BAZ BRDs with no overlapping residues at L92 in the BET BRDs (*26*). Originally, BAZ proteins were given a low predicted druggability score as well, but selective chemical probes like BAZ2-ICR have demonstrated that it is indeed possible to target this BRD family. The change of W81 in the BAZ binding pockets to leucine or valine side chains in the TRIM subfamily offers a possibility for selective inhibitor development (*26*). So far, the only affine chemical probe for the TRIM BRDs is a dual TRIM24-BRPF1 inhibitor, but fragment studies by Liu *et. al* revealed a unique binding mode of compound 6, which disorganised the water network inside the binding cavity and protruded deeper into the binding pocket than other benzoimidazolone inhibitors (*162*). Lastly, while the report in this thesis and the published chemical probe PFI-3 shed light on the fifth BRD of PB1 and SMARCA2/4, the other five PB1 BRDs remain interesting points for discussions as mutations in PB1 frequently occur in several types of cancer (*89*). Publications showed that histone binding of PB1 is mediated through an interplay of three or four of the six BRDs. PB1(2) and PB1(4) seem to be the driving force for histone binding, while PB1(1) and PB1(5) enhance the affinity (*163–165*). On the one hand, chemical probes for PB1(2)/PB1(4) could reveal more details of the binding nature by leaving full-length PB1 functional in a cellular context and inactivating single BRDs in a concentration-dependent manner. On the other hand, it would show how the effects of PFI-3 on cells are changed when preincubated with other PB1 inhibitors. PB1(2) has been rated as intermediate in the druggability score and is therefore a suitable target candidate (*26*).

Bromodomains in cellular context

As it is the aim of a chemical probe to validate possible drug candidates and determine the role of a domain or protein in a cellular context, it is important to discuss related results. To date, 14 BRD-targeting inhibitors have reached various stages in clinical trials. They are all BET inhibitors with different intended cancer applications. Unfortunately, preliminary results show significant dose-limiting toxicities, in case of BAY1238097 even with subtherapeutic doses. The adverse effects included headache, vomiting, diarrhoea, fatigue, anorexia, dysgeusia and thrombocytopenia. Also, only a fraction of patients responded to treatment. In case of OTX015, only seven out of 47 patients partially responded (*25*). Also, clinical trials for BET BRDs were referred to as premature, as none of the inhibitors shows specificity within the BET family and the knowledge about off-target effects are limited (*166*). The results clearly show a need for other therapies. Inhibition of compensating pathways could be one way of addressing the poor therapeutic outcomes and is discussed in the paragraphs below.

Recently, we published several examples which illustrate how chemical probes are used to study cellular processes. All experiments in this manuscript were performed by colleagues. The first example treated muscle differentiation (*12, 82*). During myogenesis, myoblasts nucleate to form myotubes/myofibers, which is the basic process that happens in muscle regeneration. BET family members BRD3 and BRD4 have recently been associated with myogenesis (*167*). To test this hypothesis, JQ1, Bromosporine, BAZ2A/B inhibitors BAZ2-ICR and GSK2801 and BRD7/9 inhibitor LP99 were applied to C2C12 myoblast cells. The cells were allowed to differentiate for 48, then stained and compared to a DMSO control (Figure 28, (*12*)). Strong inhibition of myoblast differentiation could be observed for JQ1 and the pan-BRD inhibitor Bromosporine, but not for the other BRD inhibitors, suggesting that indeed BET BRDs but not BRD of other families are involved in the process. These results were confirmed by gene profiling studies which measured up- or downregulation of genes after 12 hours. In contrast to cells incubated with GSK2801, BAZ2-ICR and LP99, cells with JQ1 and Bromosporine showed significant downregulation of anti-proliferative and anti-inflammatory genes. Consistently, gene expression changes were very similar in case of JQ1 and Bromosporine, confirming again that it is BET family BRDs and not other families which play a role in myogenesis.

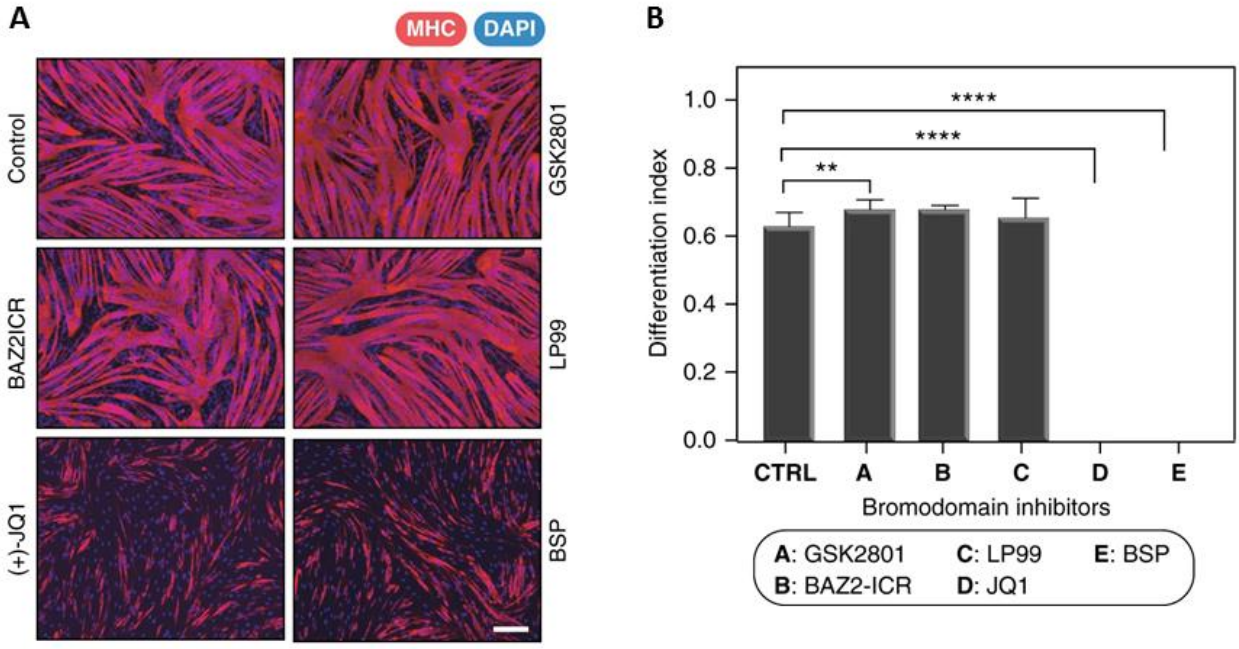


Figure 28 Muscle differentiation study **A** C1C12 myoblast cells were incubated with DMSO (control) or BRD inhibitors and allowed to differentiate for 48 h in DMEM with 2% horse serum, 10 µg/ml insulin and 10 µg/ml transferrin. Subsequently, the cells were prepared for immunofluorescence measurement with α-myosin heavy-chain antibody (red). Nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI, blue). Depicted images were taken at ×10 magnification on a Zeiss Axio Observer Z1 microscope **B** Quantification of differentiated cells after 48 h. The differentiation index was calculated by dividing the number of nuclei in myosin heavy-chain-expressing myotubes by the total number of nuclei per field. For JQ1 and BSP, significant differentiation was detected [****$p < 0.0001$]. Only minor improvement was found in GSK2801 [**$p < 0.0053$]. Incubation with BAZ2-ICR ($p = 0.1773$) and LP99 ($p = 0.1959$) had no significant effect. (Experimental data and Figure provided by Kiran Nakka and Jeffrey Dilworth, adapted from (*12*), co-author agreement obtained)

Furthermore, in the same manuscript we investigated how epigenetic treatment influences the viability of triple-negative breast cancer (TNBC) cells. TNBC patients have high recurrence and poor survival rates; there is a need for new therapies (*168*). Recent studies indicated that tumour growth and progression is downregulated by BET inhibitors (*169*). This is in line with the involvement of BRD4 in the SEC that drives transcription of oncogenes in many types of cancers (*15*). Ten different TNBC cell lines with varying metabolic states were tested with respect to viability against different BRD chemical probes. The results show indeed that in most cases BET inhibitors have anti-proliferative effects on TNBC cells whereas inhibitors targeting BRDs outside of BET had no effect (*12*). Many types of TNBC rely on an elevated glucose uptake (*170*), are marked by higher glucose transporter I (GLUT1) expression and are therefore sensitive to glucose depletion (*171*). Furthermore, higher glycolytic activity leads to specific epigenetic landscapes, because epigenetic writer substrates are derived from metabolites. For

instance, acetylation indirectly relies on glycolysis through pyruvate which ensures the availability of acetyl-CoA (*172*). To test a possible BRD inhibitor vulnerability of TNBC through glucose sensitivity, the effects of BRD inhibitors were tested in combination with BAY-876, a selective GLUT1 inhibitor (*173*). Upon BAY-876 application, decreased acetyl-CoA levels, changed NAD^+/NADH ratio and globally reduced histone H3 acetylations were detected. Interestingly, histone H4 was not reduced. Next, three of the TNBC cell lines with different metabolic rates were utilised to assess sensitivity to BRD inhibition after BAY-876 treatment. While no changes in viability were observed at 3 µM BAY-876 alone, significant decreases in viability were visible in combination with PFI-1, OF-1, or I-BRD9 (3 µM each). Among these, pan-BRPF inhibitor OF-1 was the only inhibitor that reduced viability in all three cell lines (*12*).

BRPF inhibitors were also subject of a study targeting bone osteogenesis (*82*). BRPF proteins act as scaffolding proteins and mediate contact between unmethylated histone tails and HATs of the MOZ, Ybf2/Sas3, SAS2 and Tip60 (MYST) family (*174*). Four chemical probes are available for the BRPF BRDs: the pan-BRPF chemical probes OF-1 and NI-57 and BRPF1B-selective probes GSK6853 and PFI-4 (*83, 82*). Upon application of OF-1, NI-57 and PFI-4, differentiation of primary murine bone marrow cells and human primary monocytes into bone resorbing osteoclasts was compromised. The study showed that the function of these bromodomains was necessary drive osteogenic transcriptional programs in primary cells (*82*).

The importance of bromodomains in physiological contexts is increasingly discussed also in topics which do not directly involve chemical probes. Resveratrol, a well-known polyphenolic substance with anti-oxidative effects was recently found to bind BET BRDs with micromolar affinity (*175*). This is particularly interesting, because Resveratrol was shown to induce silent mating type information regulation-1 (*SIRT1*) gene expression and downregulate NF-κB activation and therefore exerts anti-inflammatory function (*176*). Given that BRD4 is known to activate NF-κB by binding to acetylated RelA (*69, 177*), this finding underlines the effects of nutrients on the epigenetic setup. In summary, it is important to consider different metabolic states when dealing with the effects of epigenetic inhibitors.

Dual inhibitors involving bromodomains

Our study in the second part of this thesis described the rational inhibitor design for a dual BRD4-ALK inhibitor (*16*). It was prompted by a study from Ciceri *et. al*, who exposed several unintentional bromodomain inhibition effects in known kinase inhibitors (*14*). Generally, a dual inhibition effect is believed to prolong the effect of inhibitors because related pathways compensating their effects are

captured as well. The model inhibitor in our study was the PLK1-BRD4 inhibitor BI-2536 whose selectivity was changed to ALK and BRD4 by perturbing the molecule systematically on four positions (*16*).

Although the potential of dual BRD-kinase inhibitors has been recognised, only few reports include rationally designed and potent inhibitors. Dittmann *et. al* identified the phosphoinositide 3-kinase (PI3K) inhibitor LY303511 as dual antagonist that also inhibits BET BRDs (*178*), confirming that PI3K-BRD4 cross-activity is not an exception, since also PI3K-mTOR inhibitors GSK2636771 and PP-242 were found to bind BET BRDs potently (*14*). Andrews *et. al* also reported SF2535 as potent and selective PI3K-BRD4 inhibitor which stably reduces the MYCN expression levels and inhibits lock tumour growth and metastasis (*179*). The dual inhibition of PLK1 and BRD4 shown in BI-2536 promised new therapeutic approaches due to their involvement in progression of AML. Chen *et. al* performed SAR studies on BI-2536 and proposed that the dual activity could be further enhanced by chemical modifications (*180*). Hu *et. al* tuned the dual inhibitor activity of BI-2536 by improving BET affinity and diminishing PLK1 inhibition (*181*). This is an interesting attempt but hardly compares to changing one of the two targets rationally to another. Liu *et. al* made another attempt to tune PLK1-BRD4 dual activity, but compared to BI-2536 the PLK1 affinity was at least 40 times lowered (*182*).

More possibilities for dual BET/kinase inhibitors were highlighted by other recent publications. A dual approach for Aurora kinase A and BET BRDs was proposed by Felgenhauer *et. al* for high risk Neuroblastoma patients, providing an alternative target set compared to our ALK-BRD4 approach. Following the rationale of lowering MYC expression levels by BET inhibition and post-translationally destabilising MYC by inhibiting Aurora kinase A, they could prove a synergistic effect of JQ1 and Aurora kinase A inhibitor alisertib (*183*). Zhu *et. al* proposed combined inhibition of BET BRDs and MAP kinase kinase 1 (MAP2K1) as a therapy for anaplastic thyroid cancer. Using *in vitro* cell studies and mouse xenograft studies, they could show complete stop of proliferation and over 90% inhibition of tumour growth as opposed to single inhibitor use (*184*). Another combination therapy to treat myeloproliferative neoplasms (MPN) was published by Kleppe *et. al*. Despite knowledge about the importance of the JAK/STAT pathway in MPN progenitor cells, JAK inhibition is frequently accompanied by altered gene transcription, among which NF-κB signalling pathway is the most prominently activated. JQ1 was able to attenuate the inflammation pathway and therefore increased the efficacy of JAK inhibition in MPN (*185*).

Considering the mere number of combinations, BRD/kinase dual inhibition certainly promises to cover more malignancies than any other combination of protein classes. Nevertheless, BRD/HDAC dual inhibitors represent another approach considered to be effective against certain diseases. From a biochemical point of view, HDAC activity is linked to inactivation of genes, because the positive charge from histone lysine tails strengthens the interaction between histones and DNA (*186*). However, similar to BRDs, this view is too simplistic in a process that is influenced by so many parameters. In fact, HDAC activity is equally linked to activation and inactivation of genes. For instance, deacetylation by HDACs also leads to attenuated transcription of genetic repressors, so that their target genes are activated at last (*187*). Also, acetylation is clustered at promoter regions which distinguishes them from the rest of the gene bodies (*188*). The interplay between writers, erasers and readers and the epigenetic landscape altogether result in a specific genetic status. There are several malignancies where HDACs and BRDs are both involved in progression of the disease. Reports for synergistic effects of BET and HDAC inhibitors in tumour cells were shown for different types of cancers. Co-administration of panobinostat and JQ1 led to growth arrest or apoptosis in neuroblastoma and AML cell lines with more than the combined effects of both drugs. No such effects were observed in healthy cell lines (*189, 190*). The combined use of panobinostat and I-BET151 was proven to induce apoptosis in melanoma cell lines, and panobinostat with OTX-015 were successfully applied to stop proliferation of glioblastoma cells (*191, 192*). Furthermore, the use of real HDAC/BET dual inhibitors, basically consisting of known acetyl-lysine mimetics and hydroxamic acid groups, led to positive effects in leukemia cell lines *in vitro (193–195)*. None of the mentioned was tested *in vivo*.

Combined approaches between inhibitors for two epigenetic reader domains should be further invested in. The BRD-containing protein CBP has been connected to progression of AML through interaction of the oncogene c-Myb. Although a correlation with its BRD reading function has not been mentioned, it becomes evident that a combined approach for CBP/BRD4 antagonists might be beneficial (*196*). Since off-target effects between BET and CBP BRDs are common (*12*), it would be valuable to compare the effect of a BET inhibitor with CBP off-target effect against a highly specific BET inhibitor with respect to c-Myb involvement in the context of AML. Finally, Wan *et. al* reported that MOLM-13 cells were sensitised to JQ1 treatment if ENL was depleted (*19*). This is particularly interesting, because ENL/AF9 and BRD4 all reside in the super elongation complex and might therefore cooperate to drive oncogene expression in AML. This opens combination studies using BET inhibitors with SGC-iMLLT and possibly synthesis trials for BRD/YEATS dual inhibitors.

Chemical probes for YEATS domains

YEATS domains are a newly discovered family of epigenetic reader domains which have specificity for acetyl-lysines and lysines with larger modifications derived from fatty acid metabolites. Their importance in many cancer types has been emphasised by different sources and the need for chemical probes has been stated. Research done in this thesis yielded the first co-crystal structure between a YEATS domain and an inhibitory small molecule and the first human apo YEATS domain structure, thereby elucidating inherent structural features of the binding pocket such as flexibility of amino acids and conserved water molecules.

Properties of potent ENL/AF9 YEATS inhibitors

An important question for a new class of inhibitors is which chemical scaffold can mimic the natural interaction. In our case, we observed ENL/AF9, so the interaction of interest was between acetyl-/crotonyl-lysine in the ligand and the backbone of loop 6, the S58 side chain and the π-π-π with the aromatic triad F28-F59-Y78 in ENL (cf. Figure 7). We answered this question by choosing a structure-based approach and different chemical scaffolds from a fragment library. We identified an amide group flanked by two aromatic residues as the minimum requirement for potent inhibitors. Another condition that was consistent through every structure was that one of the flanking aromatic groups stacked with the H56 side chain in the front pocket. Given the discovered flexibility of the Y78, this interaction may be a must-have feature and may even be more important for potent inhibitors than the interaction with the rear pocket. Furthermore, the aromatic moiety stacked with H56 is preferably attached to the amide bond without linker atom. This theory is well-illustrated by inhibitory fragments C19, C25 and C26 (cf. Figures 23 and 27, pages 93 and 100). C25 and C26 consist of a urea-based core with one flanking aromatic moiety and one flanking saturated ring. We described this binding mode as “flipped binding mode” in the results section, because one would expect the aromatic group to face towards the rear pocket sandwiched into the aromatic triad. However, the aromatic substitution was stacked with H56 and the saturated ring faced towards the rear pocket. C19 adds another canonical condition. It contains two aromatic residues, one without linker atom at R1 (carbonyl side of amide) and with linker atom at R2 (nitrogen side of amine). In C19, the benzimidazole residue, attached without linker at R1, and not the chlorobenzyl substituent was faced towards the front pocket, which might be due to a more optimal distance to H56 and possibly due to higher importance of lower degrees of freedom in this area of the

binding pocket. The chlorobenzyl residue facing out of the binding pocket may be explained by steric hindrance of the chlorine atom.

So, first, C25 and C26 showed the general necessity of an aromatic residue stacking with H56. Considering this, one might as well name their binding modes as canonical. Second, C19 showed the preferred direct attachment of the H56 stacking aromatic moiety to the amide core. Hence, removing the spacer atom of C25 and C26 in the front pocket should in theory increase affinity. On top of these requirements, affinity can be enhanced to sub-micromolar levels if π-π interactions in the rear pocket are possible by an aromatic residue, preferably heterocyclic and decorated, so that additional hydrogen bonds with loop 6 and the side chain of E75 are formed. This is proven by the high affinity of SGC-iMLLT (Figure 26, page 98).

In summary, the properties of ENL YEATS inhibitors are ranked by importance as follows: 1) an amide-like moiety as acetyl-lysine mimetic, connecting S58 and the loop 6 backbone in a beta sheet manner, 2) an aromatic group stacking H56 – affinity is increased if directly attached to the core amide and electron-rich. On top of these two must-have properties, sub-micromolar affinity can be reached if 3) an aromatic group enabling π-π stacking faces towards the aromatic triad in the rear pocket, preferably no halogenic substituents or other sterically demanding substituents and 4) more hydrogen bond donors are placed in the rear pocket. Exceptions in which secondary amines substitute the amide-like moiety in the core have been reported, but their affinity stays in the weak micromolar range (*23*).

Selectivity in human YEATS domains

Although all human YEATS domain-containing proteins ENL, AF9, YEATS2 and GAS41 are involved in pathogenesis, so far, the only available chemical probe is SGC-iMLLT. One reason might be the flexible binding pocket, which is formed exclusively by loops and the shallow binding pocket. Judging from the binding pocket residues, ENL and AF9 can be grouped as their binding pockets are identical and also their full-length proteins reside in the super elongation complex (Figure 29). Intriguingly, the affinity of SGC-iMLLT towards ENL and AF9 differed slightly. This effect was most likely caused by residues in the periphery which can change the physicochemical parameters of the binding pocket residues. For instance, E75 could be influenced to have a slightly change pK_a value, altering the strength of the hydrogen bond with SGC-iMLLT. Such effects are unlikely to be detected by crystallography alone and are might be explained in combination with bioinformatic tools. Certainly, to elucidate the role that ENL and AF9 have physiologically, researchers would profit from a selective chemical probe, but this might

be similarly difficult to achieve as BET BRD-selective inhibitors. From a medicinal point of view, an ENL- or AF9-specific drug is unlikely to be beneficial, because ENL has been shown to replace the function of missing AF9 in cellular context (*19*). Furthermore, the dual inhibitor discussion in the upper paragraphs highlighted in many examples that multitarget drugs might even be necessary to see effects in patients.

YEATS domain research perspective

Despite the recent success in developing the first chemical probe for ENL/AF9, there is still much potential for research in this field. In my opinion, two areas are most important investigating: 1) developing chemical probes for YEATS2 and GAS41 and 2) testing SGC-iMLLT in different cellular conditions and in combination with other AML-related drugs.

As mentioned above, the binding pockets of ENL and AF9 YEATS domains are highly similar, so gaining selectivity within these two domains is very challenging if not impossible. On the other hand, both YEATS2 and GAS41 have been associated with a number of diseases, including NSCLC, glioma, colorectal and pancreatic cancer (*10, 97, 100–102*). A structural alignment of the four human YEATS domains helps defining the most significant differences between the four human YEATS domains (Figure 29). Most importantly, all four domains share histidine and a serine residue equivalent to H56 and S58 in ENL, which makes a central amide bond and a directly attached aromatic group in the front pocket a promising start for YEATS2 and GAS41 inhibitors. Interactions in the rear pocket are most likely to influence the selectivity towards other YEATS domains. Y78 of ENL is exchanged to a histidine residue in YEATS2 and GAS41, so experimenting with other heterocycles than benzimidazole could be an option. YEATS2 contains an aromatic dyad instead of a triad and an arginine residue at the position of E75 of ENL. This opens the pocket even further for bulkier heterocycles and acidic decorations may lead to interactions with arginine. GAS41 features a histidine at position F28 of ENL, which might in turn favour other heterocycles than YEATS2 with less tolerance for bulkier residues.

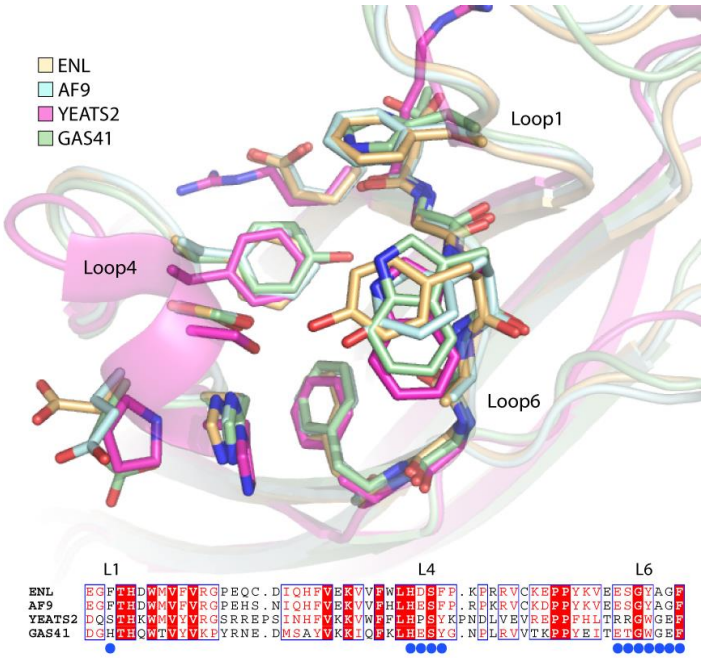


Figure 29 Structural and sequence alignment of all four human YEATS domains. Key residues are represented in stick representation and marked with blue spots under the sequence alignment. PDB codes: 5J9S, 4TMP, 5IQL, 5XTZ (all peptide-bound) (*19, 7, 18, 92*). Sequence alignment performed with ClustalO, illustrated with ESPript 3 (*61*).

The second relevant type of experiments for future research is extended testing of SGC-iMLLT in different cellular contexts and in combination with other inhibitors. YEATS domains have been implicated to drive transcriptional programs upon starvation (*46, 45, 47*). Knockout experiments with CRISPR showed downregulation of four genes (MYC, myeloperoxidase, dendrin and cathepsin G) and upregulation of CD86 in MV4;11 AML cells (*19*). The trends could be verified by application of 1 µM SGC-iMLLT to MV4;11 cells for three of the five genes (MYC, dendrin and CD86) (*24*). However, apart from CD86, the effects were much less pronounced compared to cells which were treated with 50 nM JQ1 which suggests that these genes only partially depend on YEATS domains or that their function is compensated by other genetic regulators. This is why SGC-iMLLT should be used for cells with different levels of metabolites, similar to the approach that lead to the discovery of synergy between OF-1 and BAY-876 in TNBC cells (*12*). AML is a clinically very heterogeneous disease and there are more model cell lines available to be tested. Accordingly, combination of SGC-iMLLT with different BRD inhibitors and drugs targeting PLK, FLT3, DOT1L, DNMTs and HDACs is worth investigating (*197*). Tests in MOLM-13 cells showed that the modified peptide XL-13m in combination with DOT1L inhibitor EPZ-5676 act synergistically to suppress MYC and HOXA9 (*198*). The connection between ENL/AF9 and AML is evident and future research should focus on finding the right subtype and eliminating compensating pathways.

Conclusion

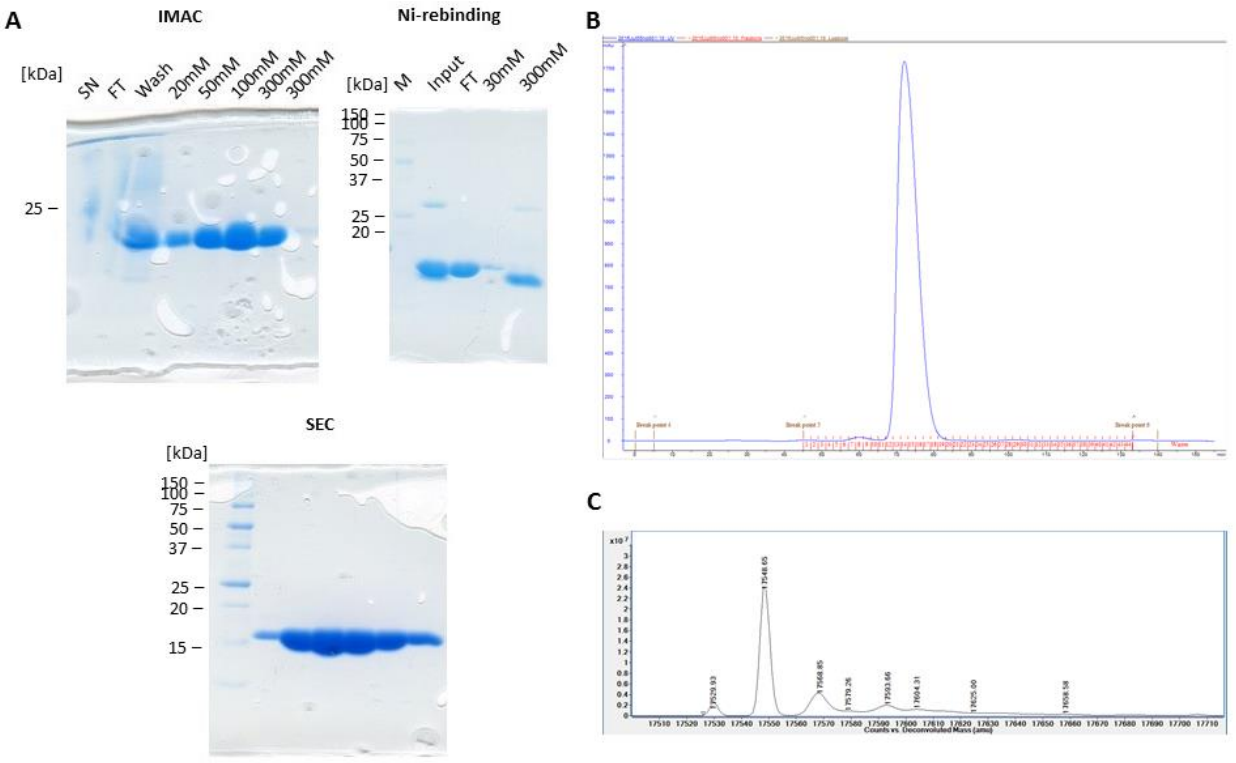
The aim of this thesis was to advance the rational development of inhibitors for histone acetyl-lysine reader domains. Three major fields have been successfully treated. Firstly, after the great success of the BET chemical probe JQ1 and the emerging development of several other inhibitors, the collaborative effort with Vassilios Myrianthopoulos *et. al* identified the use of unstructured water molecules in the binding pocket of PB1(5)/SMARCA2/4 as a means to target BRD family VIII members. The combination of computational and biophysical method was a cost-effective and data-based approach to increase the affinity of inhibitory fragments. Secondly, we systematically developed the first dual inhibitor for BRD4 and ALK^{F1174L} as possible means to tackle difficult cases of neuroblastoma. The effective elimination of PLK affinity from the precursor molecule BI-2536 solidified the general understanding of compound selectivity in kinases and opened the door to other rationally designed dual inhibitors involving BRDs. Thirdly, the research in this thesis highly contributed to the understanding of the recently emerged YEATS domain-containing proteins. This included the first YEATS apo-structure, illustrating the conservation of two water molecules in the binding pocket and inherent flexibility of Y87, the definition of requirements for potent ENL YEATS inhibitors and contribution to the development of the first YEATS chemical probe for ENL/AF9. The inhibitor studies were accompanied by the first co-crystal structures of a YEATS domain with small molecules, supporting the community with a basis to develop inhibitors YEATS2/GAS41 YEATS domains and fostering further studies on the involvement of YEATS domains in human disease.

Supplementary Information

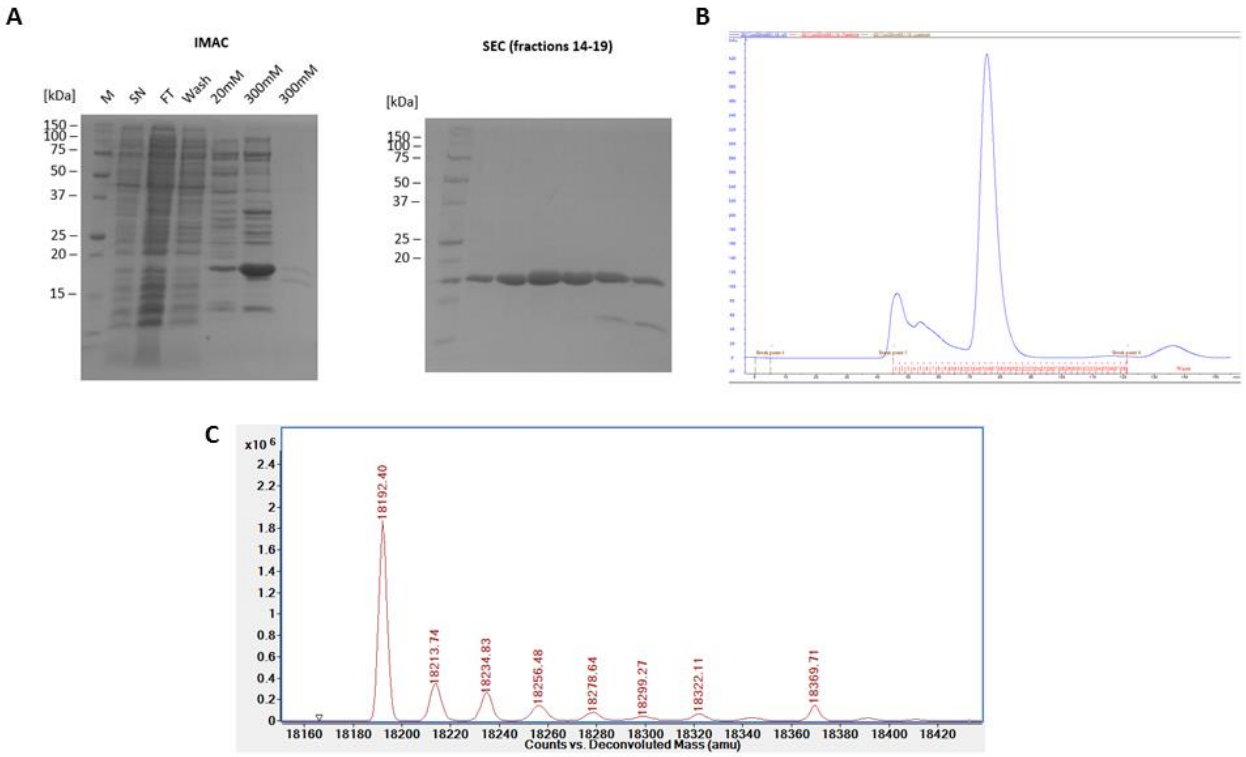
Supplementary Table S1 Protein constructs used for the contents of this thesis. Proteins marked with an asterisk were expressed and purified by colleagues LA Dutra and F Preuß and used for DSF measurements by me. Respective DNA was obtained from lab inventory except the ENL construct, which was subcloned into pNIC-CH following (*19*) by me. Constructs used for crystallisation are underlined.

Protein	Start	Stop	Vector	Antibiotic	Mass	Sequence
AF9	M1	A138	pNIC28-Bsa4	Kanamycin	18849.6	MHHHHHHSSGVDLGTENLYFQSMASSCAVQVKLELGHRAQVRKKPTVEGFTHDWMVFVRGPEHSNIQHFVEKVVFHLHESFPRPKRVCKDPPYKVEESGYAGFILPIEVYFKNKEEPRKVRFDYDLFLHLEGHPPVNHLRCEKLTFNNPTEDFRRKLLKA
BRD1	E556	A688	pNIC28-Bsa4	Kanamycin	18105.6	MHHHHHHSSGVDLGTENLYFQSMEQVAMELRLTPLTVLLRSVLDQLQDKDPARIFAQPVSLKEVPDYLDHIKHPMDFATMRKRLEAQGYKNLHEFEEDFDLIIDNCMKYNARDTVFYRAAVRLRDQGGVVLRQARREVDSIGLEEASGMHLPERPA
BRD2(1)	K71	N194	pNIC28-Bsa4	Kanamycin	17363.2	MHHHHHHSSGVDLGTENLYFQSMKPGRVTNQLQYLHKVVMKALWKHQFAWPFRQPVDAVKLGLPDYHKIIKQPMDMGTIKRRLENNYYWAASECMQDFNTMFTNCYIYNKPTDDIVLMAQTLEKIFLQKVASMPQEEQELVVTIPKN
BRD2(2)	E348	D455	pNIC28-Bsa4	Kanamycin	15431.6	MHHHHHHSSGVDLGTENLYFQSMEQLKHCNGILKELLSKKHAAYAWPFYKPVDASALGLHDYHDIIKHPMDLSTVKRKMENRDYRDAQEFAADVRLMFSNCYKYNPPDHDVVAMARKLQDVFEFRYAKMPD
BRD3(1)*	P24	E144	pNIC28-Bsa4	Kanamycin	17036.6	MHHHHHHSSGVDLGTENLYFQSMPEVSNPSKPGRKTNQLQYMQNVVVKTLWKHQFAWPFYQPVDAIKLNLPDYHKIIKNPMDMGTIKKRLENNYYWSASECMQDFNTMFTNCYIYNKPTDDIVLMAQALEKIFLQKVAQMPQEE
BRD3(2)*	G306	P416	pNIC28-Bsa4	Kanamycin	15750.1	MHHHHHHSSGVDLGTENLYFQSMGKLSEHLRYCDSILREMLSKKHAAYAWPFYKPVDAEALELHDYHDIIKHPMDLSTVKRKMDGREYPDAQGFAADVRLMFSNCYKYNPPDHEVVAMARKLQDVFEMRFAKMP
BRD4(1)	N44	E168	pNIC28-Bsa4	Kanamycin	17549.1	MHHHHHHSSGVDLGTENLYFQSMNPPPPETSNPNKPKRQTNQLQYLLRVVLKTLWKHQFAWPFQQPVDAVKLNLPDYYKIIKTPMDMGTIKKRLENNYYWNAQECIQDFNTMFTNCYIYNKPGDDIVLMAEALEKLFLQKINELPTEE
BRD4(1) no tag	N44	E168	pNIC28-Bsa4	Kanamycin	15083.5	SMNPPPPETSNPNKPKRQTNQLQYLLRVVLKTLWKHQFAWPFQQPVDAVKLNLPDYYKIIKTPMDMGTIKKRLENNYYWNAQECIQDFNTMFTNCYIYNKPGDDIVLMAEALEKLFLQKINELPTEE
BRD4(2)	K333	E460	pNIC28-Bsa4	Kanamycin	17502.0	MHHHHHHSSGVDLGTENLYFQSMKDVPDSQQHPAPEKSSKVSEQLKCCSGILKEMFAKKHAAYAWPFYKPVDVEALGLHDYCDIIKHPMDMSTIKSKLEAREYRDAQEFGADVRLMFSNCYKYNPPDHEVVAMARKLQDVFEMRFAKMPDE
BRDT(1)*	N21	E137	pNIC28-Bsa4	Kanamycin	16614.2	MHHHHHHSSGVDLGTENLYFQSMNTKKNGRLTNQLQYLQKVVLKDLWKHSFSWPFQRPVDAVKLQLPDYYTIIKNPMDLNTIKKRLENKYYAKASECIEDFNTMFSNCYLYNKPGDDIVLMAQALEKLFMQKLSQMPQEE
BRDT(2) no tag*	T266	E379	pGTVL2	Kanamycin	13746.9	SMTVKVTEQLRHCSEILKEMLAKKHFSYAWPFYNPVDVNALGLHNYYDVVKNPMDLGTIKEKMDNQEYKDAYKFAADVRLMFMNCYKYNPPDHEVVTMARMLQDVFETHFSKIPIE
BRPF1	M626	G746	pNIC28-Bsa4	Kanamycin	16839.1	MHHHHHHSSGVDLGTENLYFQSMEMQLTPFLILLRKTLEQLQEKDTGNIFSEPVPLSEVTELDEVPDYLDHIKKPMDFFTMKQNLEAYRYLNFDDFEEDFNLIVSNCLKYNAKDTIFYRAAVRLREQGGAVLRQARRQAEKMG
ENL	M1	A148	pNIC-CH	Kanamycin	18204.1	MDNQCTVQVRLELGHRAQLRKKPTTEGFTHDWMVFVRGPEQCDIQHFVEKVVFWLHDSFPKPRRVCKEPPYKVEESGYAGFIMPIEVHFKNKEEPRKVCFTYDLFLNLEGNPPVNHLRCEKLTFNNPTTEFRYKLLRAGGXMVMPEGAHHHHHH
EP300*	I1048	G1161	pNIC28-Bsa4	Kanamycin	16279.6	MHHHHHHSSGVDLGTENLYFQSMIFKPEELRQALMPTLEALYRQDPESLPFRQPVDPQLLGIPDYFDIVKSPMDLSTIKRKLDTGQYQEPWQYVDDIWLMFNNAWLYNRKTSRVYKYCSKLSEVFEQEIDPVMQSLG
GAS41	S12	I227	pET28a	Kanamycin	27414.3	MGSHHHHHHSSGTENLYFQSGGRVKGVTIVKPIVYGNVARYFGKKREEDGHTHQWTVYVKPYRNEDMSAYVKKIQFKLHESYGNPLRVVTKPPYEITETGWGEFEIIIKIFFIDPNERPVTLYHLLKLFQSDTNAMLGKKTVVSEFYDEMIFQDPTAMMQQLLTTSRQLTLGAYKHETEFAELEVKTREKLEAAKKKTSFEIAELKERLKASRETINCLKNEIRKLEEDDQAKDI

PB1(1)	H23	E156	pNIC28-Bsa4	Kanamycin	18609.1	MHHHHHHSSGVDLGTENLYFQSMHHSVSTPGPSRKRRRLSNLPTVDPIA VCHELYNTIRDYKDEQGRLLCELFIRAPKRRNQPDYYEVVSQPIDLMKIQQ KLKMEEYDDVNLLTADFQLLFNNAKSYYKPDSPEYKAACKLWDLYLRTRN EFVQKGE
PB1(2)	S178	E291	pNIC28-Bsa4	Kanamycin	15678.0	MHHHHHHSSGVDLGTENLYFQSMSPAYLKEILEQLLEAIVVATNPSGRLIS ELFQKLPSKVQYPDYYAIIKEPIDLKTIAQRIQNGSYKSIHAMAKDIDLLAKN AKTYNEPGSQVFKDANSIKKIFYMKKAEIEHHE
PB1(3)	Q356	E462	pNIC28-Bsa4	Kanamycin	15592.8	MHHHHHHSSGVDLGTENLYFQSMQLYDTVRSCRNNQGQLIAEPFYHLP SKKKYPDYYQQIKMPISLQQIRTKLKNQEYETLDHLECDLNLMFENAKRY NVPNSAIYKRVLKLQQVMQAKKKELARRDDIE
PB1(4)	G464	P605	pNIC-CTHF	Kanamycin	19339.0	MGDSMISSATSDTGSAKRKSKKNIRKQRMKILFNVVLEAREPGSGRRLCD LFMVKPSKKDYPDYYKIILEPMDLKIIEHNIRNDKYAGEEGMIEDMKLMFR NARHYNEEGSQVYNDAHILEKLLKEKRKELGPLPDDDDMASPAENLYFQS HHHHHHDYKDDDDK
PB1(5)	S613	D734	pNIC28-Bsa4	Kanamycin	17091.7	MHHHHHHSSGVDLGTENLYFQSMSGISPKKSKYMTPMQQKLNEVYEA VKNYTDKRGRRLSAIFLRLPSRSELPDYYLTIKKPMDMEKIRSHMMANKY QDIDSMVEDFVMMFNNACTYNEPESLIYKDALVLHKVLLETRRDLEGD
PB1(6)	N741	D885	pNIC28-Bsa4	Kanamycin	19913.5	MHHHHHHSSGVDLGTENLYFQSMNVTLLIQELIHNLFVSVMSHQDDEG RCYSDSLAEIPAVDPNFPNKPPLTFDIIRKNVENNRYRRLDLFQEHMFEVL ERARRMNRTDSEIYEDAVELQQFFIKIRDELCKNGEILLSPALSYTTKHLHN DVEKERKEKLPKEIEED
PCAF*	G715	D831	pNIC28-Bsa4	Kanamycin	16616.1	MHHHHHHSSGVDLGTENLYFQSMGKEKSKEPRDPDQLYSTLKSILQQVK SHQSAWPFMEPVKRTEAPGYYEVIRFPMDLKTMSERLKNRYYVSKKLFM ADLQRVFTNCKEYNPPESEYYKCANILEKFFFSKIKEAGLID
SMARCA2A	A1373	E1511	pNIC28-Bsa4	Kanamycin	18821.4	MHHHHHHSSGVDLGTENLYFQSMAEKLSPNPPKLTKQMNAIIDTVINYK DRCNVEKVPSNSQLEIEGNSSGRQLSEVFIQLPSRKELPEYYELIRKPVDFK KIKERIRNHKYRSLGDLEKDVMLLCHNAQTFNLEGSQIYEDSIVLQSVFKSA RQKIAKEEE
SMARCA2B	A1373	E1493	pNIC28-Bsa4	Kanamycin	16823.2	MHHHHHHSSGVDLGTENLYFQSMAEKLSPNPPKLTKQMNAIIDTVINYK DSSGRQLSEVFIQLPSRKELPEYYELIRKPVDFKKIKERIRNHKYRSLGDLEK DVMLLCHNAQTFNLEGSQIYEDSIVLQSVFKSARQKIAKEEE
SMARCA4A	L1451	D1569	pNIC28-Bsa4	Kanamycin	16673.1	MHHHHHHSSGVDLGTENLYFQSMLSPNPPNLTKKMKKIVDAVIKYKDSS SGRQLSEVFIQLPSRKELPEYYELIRKPVDFKKIKERIRNHKYRSLNDLEKDV MLLCQNAQTFNLEGSLIYEDSIVLQSVFTSVRQKIEKEDD
SP100*	D595	K879	pSUMO-LIC	Kanamycin	46160.6	MCSSHHHHHHGSGSGSDQEAKPSTEDLGDKKEGEYIKLKVIGQDSSEIHF KVKMTTHLKKLKESYCQRQGVPMNSLRFLFEGQRIADNHTPKELGMEEE DVIEVYQEQTGGDENINFKQSELPVTCGEVKGTLYKERFKQGTSKKCIQSE DKKWFTPREFEIEGDRGASKNWKLSIRCGGYTLKVLMENKFLPEPPSTRK KRILESHNNTLVDPCPENSNICEVCNKWGRLFCCDTCPRSFHEHCHIPSVE ANKNPWSCIFCRIKTIQERCPESQSGHQESEVLMRQMLPEEQLKCEFLLLK VYCDSKSCFFASEPYYNREGSQGPQKPMWLNKVKTSLNEQMYTRVEGF VQDMRLIFHNHKEFYREDKFTRLGIQVQDIFEKNFRNIFAIQETSK
TAF1(1)	R1398	D1524	pNIC28-Bsa4	Kanamycin	17613.1	MHHHHHHSSGVDLGTENLYFQSMRRTDPMVTLSSILESIINDMRDLPNT YPFHTPVNAKVVKDYYKIITRPMDLQTLRENVRKRLYPSREEFREHLELIVK NSATYNGPKHSLTQISQSMLDLCDEKLKEKEDKLARLEKAINPLLDDDD
TAF1(2)	D1522	D1656	pNIC28-Bsa4	Kanamycin	18331.6	MHHHHHHSSGVDLGTENLYFQSMDDDQVAFSFILDNIVTQKMMAVPD SWPFHHPVNKKFVPDYYKVIVNPMDLETIRKNISKHKYQSRESFLDDVNLI LANSVKYNGPESQYTKTAQEIVNVCYQTLTEYDEHLTQLEKDICTAKEAAL EEAELESLD
TAF1L(1)	M1401	D1522	pNIC28-Bsa4	Kanamycin	16777.1	MHHHHHHSSGVDLGTENLYFQSMVTLSSILESIINDMRDLPNTHPFHTP VNAKVVKDYYKIITRPMDLQTLRENVRKCLYPSREEFREHLELIVKNSATYN GPKHSLTQISQSMLDLCDEKLKEKEDKLARLEKAINPLLDDDD
TAF1L(2)	Q1523	D1654	pNIC28-Bsa4	Kanamycin	18000.3	MHHHHHHSSGVDLGTENLYFQSMQVAFSFILDNIVTQKMMAVPDSWP FHHPVNKKFVPDYYKMIVNPVDLETIRKNISKHKYQSRESFLDDVNLILAN SVKYNGPESQYTKTAQEIVNICYQTITEYDEHLTQLEKDICTAKEAALEEAE LESLD
TRIM33B*	D882	P1073	pNIC28-Bsa4	Kanamycin	24710.2	MHHHHHHSSGVDLGTENLYFQSMDDDPNEDWCAVCQNGGDLLCCEK CPKVFHLTCHVPTLLSFPSGDWICTFCRDIGKPEVEYDCDNLQHSKKGKTA QGLSPVDQRKCERLLLYLYCHELSIEFQEPVPASIPNYYKIIKKPMDLSTVKK KLQKKHSQHYQIPDDFVADVRLIFKNCERFNEADSEVAQAGKAVALYFED KLTEIYSDRTFAPLP
YEATS2	S202	E345	pET28a	Kanamycin	19233.7	MGSHHHHHHSSGTENLYFQSRLFVKKTIVVGNVSKYIPPDKREENDQST HKWMVYVRGSRREPSINHFVKKVWFFLHPSYKPNDLVEVREPPFHLTRR GWGEFPVRVQVHFKDSQNKRIDIIHNLKLDRTYTGLQTLGAETVVDVELH RHSLGEDCIYPQSSE



Supplementary Figure S1 Purification of BRD4(1) **A** Stained SDS-PAGE gels showing samples from IMAC, rebinding after cleaving the His-tag with TEV protease and pooled fractions after SEC **B** SEC chromatogram of FT fraction in the rebinding step **C** Deconvoluted mass of purified BRD4(1) fractions showing clean protein with sodium adducts.



Supplementary Figure S2 Purification of ENL YEATS **A** Stained SDS-PAGE gels showing samples from IMAC pooled fractions after SEC **B** SEC chromatogram of first 300 mM imidazole fraction during IMAC **C** Deconvoluted mass of purified ENL YEATS fractions showing clean protein with sodium adducts.

Supplementary Table S2 Composition of the HCS3 coarse screen

HCS3-A01	30% MPD -- 0.02M calcium chloride -- 0.1M acetate pH 4.5
HCS3-A02	0.4M sodium/potassium tartrate
HCS3-A03	0.4M ammonium phosphate monobasic
HCS3-A04	2M ammonium sulfate -- 0.1M tris pH 8.5
HCS3-A05	30% MPD -- 0.2M sodium citrate tribasic -- 0.1M HEPES pH 7.5
HCS3-A06	30% PEG4000 -- 0.2M magnesium chloride -- 0.1M tris pH 8.5
HCS3-A07	1.4M sodium acetate -- 0.1M cacodylate pH 6.5
HCS3-A08	30% 2-propanol -- 0.2M sodium citrate tribasic -- 0.1M cacodylate pH 6.5
HCS3-A09	30% PEG4000 -- 0.2M ammonium acetate -- 0.1M citrate pH 5.5
HCS3-A10	30% PEG4000 -- 0.2M ammonium acetate -- 0.1M acetate pH 4.5
HCS3-A11	0.8M ammonium phosphate monobasic -- 0.8M ammonium phosphate dibasic
HCS3-A12	30% 2-propanol -- 0.2M magnesium chloride -- 0.1M HEPES pH 7.5
HCS3-B01	30% PEG400 -- 0.2M sodium citrate tribasic -- 0.1M tris pH 8.5
HCS3-B02	28% PEG400 -- 0.2M calcium chloride -- 0.1M HEPES pH 7.5
HCS3-B03	0.2M ammonium sulfate -- 30% PEG8000 -- 0.1M cacodylate pH 6.5
HCS3-B04	1.5M lithium sulfate -- 0.1M HEPES pH 7.5
HCS3-B05	30% PEG4000 -- 0.2M lithium sulfate -- 0.1M tris pH 8.5
HCS3-B06	20% PEG8000 -- 0.2M magnesium acetate -- 0.1M cacodylate pH 6.5
HCS3-B07	30% 2-propanol -- 0.2M ammonium acetate -- 0.1M tris pH 8.5
HCS3-B08	0.2M ammonium sulfate -- 25% PEG4000 -- 0.1M acetate pH 4.5
HCS3-B09	30% MPD -- 0.2M magnesium acetate -- 0.1M cacodylate pH 6.5
HCS3-B10	30% PEG4000 -- 0.2M sodium acetate -- 0.1M tris pH 8.5
HCS3-B11	30% PEG400 -- 0.2M magnesium chloride -- 0.1M HEPES pH 7.5
HCS3-B12	20% 2-propanol -- 0.2M calcium chloride -- 0.1M acetate pH 4.5
HCS3-C01	1M sodium acetate -- 0.1M cacodylate pH 6.5
HCS3-C02	30% MPD -- 0.2M ammonium acetate -- 0.1M citrate pH 5.5
HCS3-C03	20% 2-propanol -- 0.2M sodium citrate tribasic -- 0.1M HEPES pH 7.5
HCS3-C04	0.2M sodium acetate -- 30% PEG8000 -- 0.1M cacodylate pH 6.5
HCS3-C05	0.8M sodium/potassium tartrate -- 0.1M HEPES pH 7.5
HCS3-C06	0.2M ammonium sulfate -- 30% PEG8000
HCS3-C07	0.2M ammonium sulfate -- 30% PEG4000
HCS3-C08	2M ammonium sulfate
HCS3-C09	4M sodium formate
HCS3-C10	2M sodium formate -- 0.1M acetate pH 4.5
HCS3-C11	0.4M Na phosphate monobasic -- 0.4M K phosphate monobasic -- 0.1M HEPES pH 7.5
HCS3-C12	8% PEG8000 -- 0.1M tris pH 8.5
HCS3-D01	8% PEG4000 -- 0.1M acetate pH 4.5
HCS3-D02	1.4M sodium citrate tribasic -- 0.1M HEPES pH 7.5
HCS3-D03	2M ammonium sulfate -- 2% PEG400 -- 0.1M HEPES pH 7.5
HCS3-D04	20% PEG4000 -- 20% 2-propanol -- 0.1M citrate pH 5.5
HCS3-D05	20% PEG4000 -- 10% 2-propanol -- 0.1M HEPES pH 7.5
HCS3-D06	20% PEG8000 -- 0.04M potassium phosphate monobasic
HCS3-D07	30% PEG1500
HCS3-D08	0.3M magnesium formate
HCS3-D09	18% PEG8000 -- 0.2M zinc acetate -- 0.1M cacodylate pH 6.5
HCS3-D10	18% PEG8000 -- 0.2M calcium acetate -- 0.1M cacodylate pH 6.5
HCS3-D11	2M ammonium sulfate -- 0.1M acetate pH 4.5
HCS3-D12	0.8M ammonium phosphate monobasic -- 0.1M tris pH 8.5
HCS3-E01	10% PEG6000 -- 2M sodium chloride
HCS3-E02	0.5M sodium chloride -- 0.01M magnesium chloride -- 0.01M cetrimonium bromide
HCS3-E03	25% ethylene glycol
HCS3-E04	35%(v/v) dioxane
HCS3-E05	2M ammonium sulfate -- 5% 2-propanol
HCS3-E06	8% PEG20000 -- 8% PEG500MME
HCS3-E07	10% PEG8000 -- 10% PEG1000
HCS3-E08	1.5M sodium chloride -- 10% ethanol

HCS3-E09	2M sodium chloride -- 0.1M acetate pH 4.5
HCS3-E10	30% MPD -- 0.2M sodium chloride -- 0.1M acetate pH 4.5
HCS3-E11	1M 1,6-hexanediol -- 0.01M cobalt chloride -- 0.1M acetate pH 4.5
HCS3-E12	30% PEG400 -- 0.1M cadmium chloride -- 0.1M acetate pH 4.5
HCS3-F01	0.2M ammonium sulfate -- 30% PEG2000MME -- 0.1M acetate pH 4.5
HCS3-F02	0.2M sodium/potassium tartrate -- 2M ammonium sulfate -- 0.1M citrate pH 5.5
HCS3-F03	0.5M ammonium sulfate -- 1M lithium sulfate -- 0.1M citrate pH 5.5
HCS3-F04	0.5M sodium chloride -- 2% jeffamine M-600 -- 0.1M citrate pH 5.5
HCS3-F05	35% tert-butanol -- 0.1M citrate pH 5.5
HCS3-F06	10% jeffamine M-600 -- 0.01M iron chloride -- 0.1M citrate pH 5.5
HCS3-F07	2.5M 1,6-hexanediol -- 0.1M citrate pH 5.5
HCS3-F08	1.6M magnesium sulfate -- 0.1M MES pH 6.5
HCS3-F09	0.1M Na phosphate monobasic -- 0.1M K phosphate dibasic -- 2M sodium chloride -- 0.1M MES pH 6.5
HCS3-F10	12% PEG20000 -- 0.1M MES pH 6.5
HCS3-F11	1.6M ammonium sulfate -- 10%(v/v) dioxane -- 0.1M MES pH 6.5
HCS3-F12	30% jeffamine M-600 -- 0.04M cesium chloride -- 0.1M MES pH 6.5
HCS3-G01	1.8M ammonium sulfate -- 0.01M cobalt chloride -- 0.1M MES pH 6.5
HCS3-G02	0.2M ammonium sulfate -- 30% PEG5000MME -- 0.1M MES pH 6.5
HCS3-G03	25% PEG500MME -- 0.01M zinc sulfate -- 0.1M MES pH 6.5
HCS3-G04	1.6M sodium citrate tribasic
HCS3-G05	30% MPD -- 0.5M ammonium sulfate -- 0.1M HEPES pH 7.5
HCS3-G06	5% MPD -- 10% PEG6000 -- 0.1M HEPES pH 7.5
HCS3-G07	20% jeffamine M-600 -- 0.1M HEPES pH 7.5
HCS3-G08	1.6M ammonium sulfate -- 0.1M sodium chloride -- 0.1M HEPES pH 7.5
HCS3-G09	2M ammonium formate -- 0.1M HEPES pH 7.5
HCS3-G10	1M sodium acetate -- 0.04M cadmium sulfate -- 0.1M HEPES pH 7.5
HCS3-G11	70% MPD -- 0.1M HEPES pH 7.5
HCS3-G12	4.2M sodium chloride -- 0.1M HEPES pH 7.5
HCS3-H01	10% PEG8000 -- 8% ethylene glycol -- 0.1M HEPES pH 7.5
HCS3-H02	20%(w/v) PEG10000 -- 0.1M HEPES pH 7.5
HCS3-H03	3.4M 1,6-hexanediol -- 0.2M magnesium chloride -- 0.1M tris pH 8.5
HCS3-H04	25% tert-butanol -- 0.1M tris pH 8.5
HCS3-H05	1M lithium sulfate -- 0.01M nickel chloride -- 0.1M tris pH 8.5
HCS3-H06	1.6M ammonium sulfate -- 12% glycerol -- 0.1M tris pH 8.5
HCS3-H07	50% MPD -- 0.2M ammonium phosphate dibasic -- 0.1M tris pH 8.5
HCS3-H08	20% ethanol -- 0.1M tris pH 8.5
HCS3-H09	20% PEG2000MME -- 0.01M nickel chloride -- 0.1M tris pH 8.5
HCS3-H10	0.1M sodium chloride -- 20% PEG500MME -- 0.1M bicine pH 9.0
HCS3-H11	2M magnesium chloride -- 0.1M bicine pH 9.0
HCS3-H12	2%(v/v) dioxane -- 10% PEG20000 -- 0.1M bicine pH 9.0

Supplementary Table S3 Composition of the HIN3 coarse screen

HIN3-A01	2M ammonium sulfate -- 0.1M citrate pH 3.5
HIN3-A02	2M ammonium sulfate -- 0.1M acetate pH 4.5
HIN3-A03	2M ammonium sulfate -- 0.1M bis-tris pH 5.5
HIN3-A04	2M ammonium sulfate -- 0.1M bis-tris pH 6.5
HIN3-A05	2M ammonium sulfate -- 0.1M HEPES pH 7.5
HIN3-A06	2M ammonium sulfate -- 0.1M tris pH 8.5
HIN3-A07	3M sodium chloride -- 0.1M citrate pH 3.5
HIN3-A08	3M sodium chloride -- 0.1M acetate pH 4.5
HIN3-A09	3M sodium chloride -- 0.1M bis-tris pH 5.5
HIN3-A10	3M sodium chloride -- 0.1M bis-tris pH 6.5
HIN3-A11	3M sodium chloride -- 0.1M HEPES pH 7.5
HIN3-A12	3M sodium chloride -- 0.1M tris pH 8.5

HIN3-B01	0.3M magnesium formate -- 0.1M bis-tris pH 5.5
HIN3-B02	0.5M magnesium formate -- 0.1M bis-tris pH 6.5
HIN3-B03	0.5M magnesium formate -- 0.1M HEPES pH 7.5
HIN3-B04	0.3M magnesium formate -- 0.1M tris pH 8.5
HIN3-B05	1.26M sodium phosphate monobasic -- 0.14M potassium phosphate dibasic
HIN3-B06	0.49M sodium phosphate monobasic -- 0.91M potassium phosphate dibasic
HIN3-B07	0.056M sodium phosphate monobasic -- 1.344M potassium phosphate dibasic
HIN3-B08	1.4M sodium citrate tribasic -- 0.1M HEPES pH 7.5
HIN3-B09	1.8M ammonium citrate
HIN3-B10	0.8M succinic acid
HIN3-B11	2.1M DL- malic acid
HIN3-B12	2.8M sodium acetate
HIN3-C01	3.5M sodium formate
HIN3-C02	1.1M ammonium tartrate
HIN3-C03	2.4M sodium malonate
HIN3-C04	35% tacsimate
HIN3-C05	60% tacsimate
HIN3-C06	1.5M ammonium sulfate -- 0.1M sodium chloride -- 0.1M bis-tris pH 6.5
HIN3-C07	0.8M sodium/potassium tartrate -- 0.5% PEG5000MME -- 0.1M tris pH 8.5
HIN3-C08	1M ammonium sulfate -- 1% PEG3350 -- 0.1M bis-tris pH 5.5
HIN3-C09	1.1M sodium malonate -- 0.5% jeffamine ED-2003 -- 0.1M HEPES pH 7.0
HIN3-C10	1M succinic acid -- 1% PEG2000MME -- 0.1M HEPES pH 7.0
HIN3-C11	1M ammonium sulfate -- 0.5% PEG8000 -- 0.1M HEPES pH 7.0
HIN3-C12	15% tacsimate -- 2% PEG3350 -- 0.1M HEPES pH 7.0
HIN3-D01	25% PEG1500
HIN3-D02	30% jeffamine M-600 -- 0.1M HEPES pH 7.0
HIN3-D03	30% jeffamine ED-2003 -- 0.1M HEPES pH 7.0
HIN3-D04	25% PEG3350 -- 0.1M citrate pH 3.5
HIN3-D05	25% PEG3350 -- 0.1M acetate pH 4.5
HIN3-D06	25% PEG3350 -- 0.1M bis-tris pH 5.5
HIN3-D07	25% PEG3350 -- 0.1M bis-tris pH 6.5
HIN3-D08	25% PEG3350 -- 0.1M HEPES pH 7.5
HIN3-D09	25% PEG3350 -- 0.1M tris pH 8.5
HIN3-D10	20% PEG5000MME -- 0.1M bis-tris pH 6.5
HIN3-D11	28% PEG2000MME -- 0.1M bis-tris pH 6.5
HIN3-D12	0.2M calcium chloride -- 45% MPD -- 0.1M bis-tris pH 5.5
HIN3-E01	0.2M calcium chloride -- 45% MPD -- 0.1M bis-tris pH 6.5
HIN3-E02	0.2M ammonium acetate -- 45% MPD -- 0.1M bis-tris pH 5.5
HIN3-E03	0.2M ammonium acetate -- 45% MPD -- 0.1M bis-tris pH 6.5
HIN3-E04	0.2M ammonium acetate -- 45% MPD -- 0.1M HEPES pH 7.5
HIN3-E05	0.2M ammonium acetate -- 45% MPD -- 0.1M tris pH 8.5
HIN3-E06	0.05M calcium chloride -- 30% PEG500MME -- 0.1M bis-tris pH 6.5
HIN3-E07	0.05M magnesium chloride -- 30% PEG500MME -- 0.1M HEPES pH 7.5
HIN3-E08	0.2M potassium chloride -- 35% pentaerythritol propoxylate 5/4 -- 0.1M HEPES pH 7.5
HIN3-E09	0.05M ammonium sulfate -- 30% pentaerythritol ethoxylate 15/4 -- 0.1M bis-tris pH 6.5
HIN3-E10	45% polypropylene glycol 400 -- 0.1M bis-tris pH 6.5
HIN3-E11	0.02M magnesium chloride -- 22% polyacrylic acid 5100 -- 0.1M HEPES pH 7.5
HIN3-E12	0.01M cobalt chloride -- 20% polyvinylpyrrolidone -- 0.1M tris pH 8.5
HIN3-F01	0.2M L-Proline -- 10% PEG3350 -- 0.1M HEPES pH 7.5
HIN3-F02	0.2M trimethylamine N-oxide -- 20% PEG2000MME -- 0.1M tris pH 8.5
HIN3-F03	5% tacsimate -- 10% PEG5000MME -- 0.1M HEPES pH 7.0
HIN3-F04	0.005M (each) MgCl2, CoCl2, Ni(II)Cl2, CdCl -- 12% PEG3350 -- 0.1M HEPES pH 7.5
HIN3-F05	0.1M ammonium acetate -- 17%(w/v) PEG10000 -- 0.1M bis-tris pH 5.5
HIN3-F06	0.2M ammonium sulfate -- 25% PEG3350 -- 0.1M bis-tris pH 5.5
HIN3-F07	0.2M ammonium sulfate -- 25% PEG3350 -- 0.1M bis-tris pH 6.5
HIN3-F08	0.2M ammonium sulfate -- 25% PEG3350 -- 0.1M HEPES pH 7.5
HIN3-F09	0.2M ammonium sulfate -- 25% PEG3350 -- 0.1M tris pH 8.5

HIN3-F10	0.2M sodium chloride -- 25% PEG3350 -- 0.1M bis-tris pH 5.5
HIN3-F11	0.2M sodium chloride -- 25% PEG3350 -- 0.1M bis-tris pH 6.5
HIN3-F12	0.2M sodium chloride -- 25% PEG3350 -- 0.1M HEPES pH 7.5
HIN3-G01	0.2M sodium chloride -- 25% PEG3350 -- 0.1M tris pH 8.5
HIN3-G02	0.2M lithium sulfate -- 25% PEG3350 -- 0.1M bis-tris pH 5.5
HIN3-G03	0.2M lithium sulfate -- 25% PEG3350 -- 0.1M bis-tris pH 6.5
HIN3-G04	0.2M lithium sulfate -- 25% PEG3350 -- 0.1M HEPES pH 7.5
HIN3-G05	0.2M lithium sulfate -- 25% PEG3350 -- 0.1M tris pH 8.5
HIN3-G06	0.2M ammonium acetate -- 25% PEG3350 -- 0.1M bis-tris pH 5.5
HIN3-G07	0.2M ammonium acetate -- 25% PEG3350 -- 0.1M bis-tris pH 6.5
HIN3-G08	0.2M ammonium acetate -- 25% PEG3350 -- 0.1M HEPES pH 7.5
HIN3-G09	0.2M ammonium acetate -- 25% PEG3350 -- 0.1M tris pH 8.5
HIN3-G10	0.2M magnesium chloride -- 25% PEG3350 -- 0.1M bis-tris pH 5.5
HIN3-G11	0.2M magnesium chloride -- 25% PEG3350 -- 0.1M bis-tris pH 6.5
HIN3-G12	0.2M magnesium chloride -- 25% PEG3350 -- 0.1M HEPES pH 7.5
HIN3-H01	0.2M magnesium chloride -- 25% PEG3350 -- 0.1M tris pH 8.5
HIN3-H02	0.2M sodium/potassium tartrate -- 20% PEG3350
HIN3-H03	0.2M sodium malonate -- 20% PEG3350
HIN3-H04	0.2M ammonium citrate -- 20% PEG3350
HIN3-H05	0.1M succinic acid -- 15% PEG3350
HIN3-H06	0.2M sodium formate -- 20% PEG3350
HIN3-H07	0.15M DL- malic acid -- 20% PEG3350
HIN3-H08	0.1M magnesium formate -- 15% PEG3350
HIN3-H09	0.005M zinc acetate -- 20% PEG3350
HIN3-H10	0.2M sodium citrate tribasic -- 20% PEG3350
HIN3-H11	0.1M potassium thiocyanate -- 30% PEG2000MME
HIN3-H12	0.1M potassium bromide -- 30% PEG2000MME

Supplementary Table S4 Composition of the LFS6 coarse screen

LFS6-A01	30% PEG1000 -- 0.1M SPG pH 6.0
LFS6-A02	30% PEG1000 -- 0.1M SPG pH 7.0
LFS6-A03	30% PEG1000 -- 0.1M SPG pH 8.0
LFS6-A04	60% MPD -- 0.1M SPG pH 6.0
LFS6-A05	60% MPD -- 0.1M SPG pH 7.0
LFS6-A06	60% MPD -- 0.1M SPG pH 8.0
LFS6-A07	20% PEG6000 -- 10% ethylene glycol -- 0.2M sodium chloride
LFS6-A08	20% PEG6000 -- 10% ethylene glycol -- 0.2M ammonium chloride
LFS6-A09	20% PEG6000 -- 10% ethylene glycol -- 0.2M lithium chloride
LFS6-A10	20% PEG6000 -- 10% ethylene glycol -- 0.1M magnesium chloride
LFS6-A11	20% PEG6000 -- 10% ethylene glycol -- 0.1M calcium chloride
LFS6-A12	20% PEG6000 -- 10% ethylene glycol -- 0.01M zinc chloride
LFS6-B01	30% PEG1000 -- 0.1M MIB pH 6.0
LFS6-B02	30% PEG1000 -- 0.1M MIB pH 7.0
LFS6-B03	30% PEG1000 -- 0.1M MIB pH 8.0
LFS6-B04	60% MPD -- 0.1M MIB pH 6.0
LFS6-B05	60% MPD -- 0.1M MIB pH 7.0
LFS6-B06	60% MPD -- 0.1M MIB pH 8.0
LFS6-B07	20% PEG6000 -- 10% ethylene glycol -- 0.1M MES pH 6.0 -- 0.2M sodium chloride
LFS6-B08	20% PEG6000 -- 10% ethylene glycol -- 0.1M MES pH 6.0 -- 0.2M ammonium chloride
LFS6-B09	20% PEG6000 -- 10% ethylene glycol -- 0.1M MES pH 6.0 -- 0.2M lithium chloride
LFS6-B10	20% PEG6000 -- 10% ethylene glycol -- 0.1M MES pH 6.0 -- 0.1M magnesium chloride
LFS6-B11	20% PEG6000 -- 10% ethylene glycol -- 0.1M MES pH 6.0 -- 0.1M calcium chloride
LFS6-B12	20% PEG6000 -- 10% ethylene glycol -- 0.1M MES pH 6.0 -- 0.01M zinc chloride
LFS6-C01	30% PEG1000 -- 0.1M PCB pH 6.0

LFS6-C02	30% PEG1000 -- 0.1M PCB pH 7.0
LFS6-C03	30% PEG1000 -- 0.1M PCB pH 8.0
LFS6-C04	60% MPD -- 0.1M PCB pH 6.0
LFS6-C05	60% MPD -- 0.1M PCB pH 7.0
LFS6-C06	60% MPD -- 0.1M PCB pH 8.0
LFS6-C07	20% PEG6000 -- 10% ethylene glycol -- 0.1M HEPES pH 7.0 -- 0.2M sodium chloride
LFS6-C08	20% PEG6000 -- 10% ethylene glycol -- 0.1M HEPES pH 7.0 -- 0.2M ammonium chloride
LFS6-C09	20% PEG6000 -- 10% ethylene glycol -- 0.1M HEPES pH 7.0 -- 0.2M lithium chloride
LFS6-C10	20% PEG6000 -- 10% ethylene glycol -- 0.1M HEPES pH 7.0 -- 0.1M magnesium chloride
LFS6-C11	20% PEG6000 -- 10% ethylene glycol -- 0.1M HEPES pH 7.0 -- 0.1M calcium chloride
LFS6-C12	20% PEG6000 -- 10% ethylene glycol -- 0.1M HEPES pH 7.0 -- 0.01M zinc chloride
LFS6-D01	30% PEG1000 -- 0.1M MMT pH 6.0
LFS6-D02	30% PEG1000 -- 0.1M MMT pH 7.0
LFS6-D03	30% PEG1000 -- 0.1M MMT pH 8.0
LFS6-D04	60% MPD -- 0.1M MMT pH 6.0
LFS6-D05	60% MPD -- 0.1M MMT pH 7.0
LFS6-D06	60% MPD -- 0.1M MMT pH 8.0
LFS6-D07	20% PEG6000 -- 10% ethylene glycol -- 0.1M tris pH 7.5 -- 0.2M sodium chloride
LFS6-D08	20% PEG6000 -- 10% ethylene glycol -- 0.1M tris pH 7.5 -- 0.2M ammonium chloride
LFS6-D09	20% PEG6000 -- 10% ethylene glycol -- 0.1M tris pH 7.5 -- 0.2M lithium chloride
LFS6-D10	20% PEG6000 -- 10% ethylene glycol -- 0.1M tris pH 7.5 -- 0.1M magnesium chloride
LFS6-D11	20% PEG6000 -- 10% ethylene glycol -- 0.1M tris pH 7.5 -- 0.1M calcium chloride
LFS6-D12	20% PEG6000 -- 10% ethylene glycol -- 0.1M tris pH 7.5 -- 0.01M zinc chloride
LFS6-E01	20% PEG3350 -- 10% ethylene glycol -- 0.2M sodium fluoride
LFS6-E02	20% PEG3350 -- 10% ethylene glycol -- 0.2M sodium bromide
LFS6-E03	20% PEG3350 -- 10% ethylene glycol -- 0.2M sodium iodide
LFS6-E04	20% PEG3350 -- 10% ethylene glycol -- 0.2M potassium thiocyanate
LFS6-E05	20% PEG3350 -- 10% ethylene glycol -- 0.2M sodium nitrate
LFS6-E06	20% PEG3350 -- 10% ethylene glycol -- 0.2M sodium formate
LFS6-E07	20% PEG3350 -- 10% ethylene glycol -- 0.2M sodium acetate
LFS6-E08	20% PEG3350 -- 10% ethylene glycol -- 0.2M sodium sulfate
LFS6-E09	20% PEG3350 -- 10% ethylene glycol -- 0.2M sodium/potassium tartrate
LFS6-E10	20% PEG3350 -- 10% ethylene glycol -- 0.2M sodium/potassium phosphate
LFS6-E11	20% PEG3350 -- 10% ethylene glycol -- 0.2M potassium citrate tribasic
LFS6-E12	20% PEG3350 -- 10% ethylene glycol -- 0.2M sodium malonate
LFS6-F01	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 6.5 -- 0.2M sodium fluoride
LFS6-F02	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 6.5 -- 0.2M sodium bromide
LFS6-F03	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 6.5 -- 0.2M sodium iodide
LFS6-F04	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 6.5 -- 0.2M potassium thiocyanate
LFS6-F05	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 6.5 -- 0.2M sodium nitrate
LFS6-F06	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 6.5 -- 0.2M sodium formate
LFS6-F07	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 6.5 -- 0.2M sodium acetate
LFS6-F08	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 6.5 -- 0.2M sodium sulfate
LFS6-F09	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 6.5 -- 0.2M sodium/potassium tartrate
LFS6-F10	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 6.5 -- 0.02M sodium/potassium phosphate
LFS6-F11	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 6.5 -- 0.2M potassium citrate tribasic
LFS6-F12	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 6.5 -- 0.2M sodium malonate
LFS6-G01	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 7.5 -- 0.2M sodium fluoride
LFS6-G02	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 7.5 -- 0.2M sodium bromide
LFS6-G03	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 7.5 -- 0.2M sodium iodide
LFS6-G04	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 7.5 -- 0.2M potassium thiocyanate
LFS6-G05	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 7.5 -- 0.2M sodium nitrate
LFS6-G06	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 7.5 -- 0.2M sodium formate
LFS6-G07	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 7.5 -- 0.2M sodium acetate
LFS6-G08	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 7.5 -- 0.2M sodium sulfate
LFS6-G09	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 7.5 -- 0.2M sodium/potassium tartrate
LFS6-G10	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 7.5 -- 0.02M sodium/potassium phosphate

LFS6-G11	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 7.5 -- 0.2M potassium citrate tribasic
LFS6-G12	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 7.5 -- 0.2M sodium malonate
LFS6-H01	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 8.5 -- 0.2M sodium fluoride
LFS6-H02	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 8.5 -- 0.2M sodium bromide
LFS6-H03	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 8.5 -- 0.2M sodium iodide
LFS6-H04	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 8.5 -- 0.2M potassium thiocyanate
LFS6-H05	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 8.5 -- 0.2M sodium nitrate
LFS6-H06	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 8.5 -- 0.2M sodium formate
LFS6-H07	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 8.5 -- 0.2M sodium acetate
LFS6-H08	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 8.5 -- 0.2M sodium sulfate
LFS6-H09	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 8.5 -- 0.2M sodium/potassium tartrate
LFS6-H10	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 8.5 -- 0.02M sodium/potassium phosphate
LFS6-H11	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 8.5 -- 0.2M potassium citrate tribasic
LFS6-H12	20% PEG3350 -- 10% ethylene glycol -- 0.1M bis-tris-propane pH 8.5 -- 0.2M sodium malonate

Supplementary Table S5 Composition of the BCS coarse screen

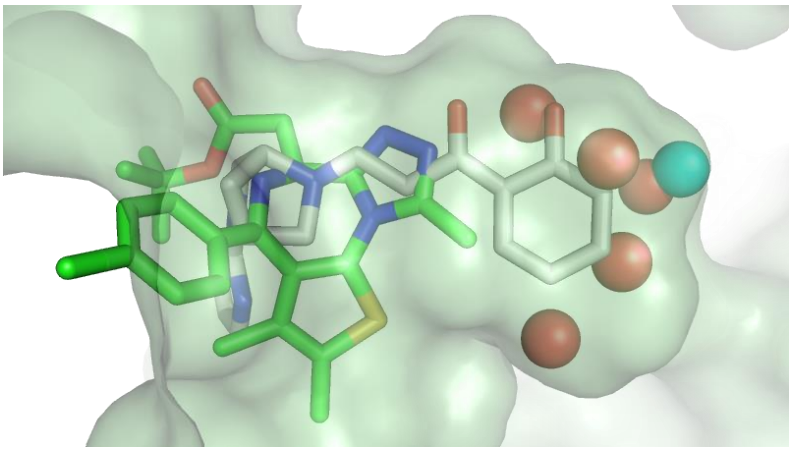
BCS-A01	30% PEG Smear Low -- 0.1M acetate pH 4.5
BCS-A02	30% PEG Smear Low -- 0.1M citrate pH 5.5
BCS-A03	30% PEG Smear Low -- 0.1M MES pH 6.5
BCS-A04	25% PEG Smear Medium -- 0.1M acetate pH 4.5
BCS-A05	25% PEG Smear Medium -- 0.1M citrate pH 5.5
BCS-A06	25% PEG Smear Medium -- 0.1M MES pH 6.5
BCS-A07	20% PEG Smear High -- 0.1M acetate pH 4.5
BCS-A08	20% PEG Smear High -- 0.1M citrate pH 5.5
BCS-A09	20% PEG Smear High -- 0.1M MES pH 6.5
BCS-A10	22% PEG Smear Broad -- 0.1M acetate pH 4.5
BCS-A11	22% PEG Smear Broad -- 0.1M citrate pH 5.5
BCS-A12	22% PEG Smear Broad -- 0.1M MES pH 6.5
BCS-B01	30% PEG Smear Low -- 0.1M HEPES pH 7.5
BCS-B02	30% PEG Smear Low -- 0.1M tris pH 8.5
BCS-B03	30% PEG Smear Low -- 0.1M bicine pH 9.0
BCS-B04	25% PEG Smear Medium -- 0.1M HEPES pH 7.5
BCS-B05	25% PEG Smear Medium -- 0.1M tris pH 8.5
BCS-B06	25% PEG Smear Medium -- 0.1M bicine pH 9.0
BCS-B07	20% PEG Smear High -- 0.1M HEPES pH 7.5
BCS-B08	20% PEG Smear High -- 0.1M tris pH 8.5
BCS-B09	20% PEG Smear High -- 0.1M bicine pH 9.0
BCS-B10	22% PEG Smear Broad -- 0.1M HEPES pH 7.5
BCS-B11	22% PEG Smear Broad -- 0.1M tris pH 8.5
BCS-B12	22% PEG Smear Broad -- 0.1M bicine pH 9.0
BCS-C01	35% PEG Smear Low
BCS-C02	28% PEG Smear Low -- 0.1M acetate pH 4.5 -- 0.2M ammonium acetate -- 5% ethylene glycol
BCS-C03	28% PEG Smear Medium -- 0.15M sodium chloride
BCS-C04	25% PEG Smear Medium -- 0.1M cacodylate pH 5.5 -- 0.2M ammonium sulfate
BCS-C05	25% PEG Smear Medium -- 0.1M citrate pH 5.5 -- 0.1M rubidium chloride -- 0.1M Na/K phosphate
BCS-C06	22.5% PEG Smear High -- 0.2M KCl
BCS-C07	15% PEG Smear High -- 0.1M citrate pH 5.5 -- 0.15M ammonium acetate
BCS-C08	28% PEG Smear Broad -- 0.05M L-Proline -- 5% glycerol
BCS-C09	20% PEG Smear Broad -- 0.1M citrate pH 5.5 -- 0.15M magnesium acetate
BCS-C10	25% PEG Smear Broad -- 0.1M acetate pH 4.5 -- 0.2M ammonium sulfate
BCS-C11	25% PEG Smear Low -- 0.1M MES pH 6.0 -- 0.2M Na/K tartrate
BCS-C12	22.5% PEG Smear Medium -- 0.1M PIPES pH 7.0 -- 0.1M CaCl2 -- 0.1M MgCl2
BCS-D01	22.5% PEG Smear Low -- 0.1M cacodylate pH 5.0 -- 0.2M ammonium nitrate
BCS-D02	22.5% PEG Smear Low -- 0.1M MES pH 6.5 -- 10% 2-propanol
BCS-D03	20% PEG Smear Medium -- 0.1M MES pH 6.0 -- 0.15M ammonium nitrate -- 5% ethylene glycol

BCS-D04	20% PEG Smear Medium -- 0.1M Na/K phosphate pH 6.2 -- 0.2M sodium formate -- 10% glycerol
BCS-D05	30% PEG Smear Medium -- 0.1M ADA pH 6.5 -- 0.2M lithium sulfate
BCS-D06	12% PEG Smear High -- 0.1M MES pH 6.5 -- 0.1M potassium thiocyanate -- 0.1M sodium bromide
BCS-D07	18% PEG Smear High -- 0.1M ADA pH 6.5-- 0.2M ammonium sulfate
BCS-D08	15% PEG Smear Broad -- 0.1M MES pH 6.0 -- 0.15M CaCl2 -- 5% glycerol
BCS-D09	15% PEG Smear Broad -- 0.1M cacodylate pH 5.0 -- 5% tacsimate -- 10% ethylene glycol
BCS-D10	28% PEG Smear Broad -- 0.1M Na/K phosphatepH 6.2 -- 0.2M sodium chloride
BCS-D11	22.5% PEG Smear Medium -- 0.1M citrate pH 5.5 -- 0.1M ammonium sulfate -- 0.05M magnesium sulfate
BCS-D12	22.5% PEG Smear Medium -- 0.1M bis-tris-propane pH 8.0 -- 0.01M Co(II)Cl2 -- 0.2M MgCl2 -- 2% glycerol
BCS-E01	25% PEG Smear Low -- 0.1M MES pH 6.5 -- 0.05M magnesium acetate -- 0.05M MgCl2
BCS-E02	16% PEG Smear Low -- 0.1M HEPES pH 7.5 -- 5% ethylene glycol -- 0.1M KCl
BCS-E03	20% PEG Smear Medium -- 0.1M bis-tris pH 7.5 -- 0.1M zinc acetate -- 0.1M zinc chloride
BCS-E04	20% PEG Smear Medium -- 0.1M PIPES pH 7.0 -- 0.1M MgCl2 -- 0.1M KCl
BCS-E05	28% PEG Smear Medium -- 0.1M HEPES pH 7.5 -- 0.05M magnesium sulfate
BCS-E06	15% PEG Smear High -- 0.1M HEPES pH 7.5 -- 0.1M Na/K phosphate-- 10% ethylene glycol
BCS-E07	25% PEG Smear High -- 0.1M PIPES pH 7.0 -- 0.1M magnesium formate -- 0.1M rubidium chloride
BCS-E08	25% PEG Smear Broad -- 0.1M HEPES pH 7.2 -- 0.2M lithium sulfate
BCS-E09	20% PEG Smear Broad -- 0.1M HEPES pH 7.5 -- 0.2M ammonium nitrate
BCS-E10	30% PEG Smear Broad -- 0.1M HEPES pH 7.5 -- 0.1M MgCl2 -- 0.1M rubidium chloride
BCS-E11	22.5% PEG Smear High -- 0.1M bis-tris-propane pH 8.0 -- 0.05M MgCl2 -- 0.05M sodium citrate tribasic
BCS-E12	22.5% PEG Smear High -- 0.1M bicine pH 9.0 -- 7.5% tacsimate -- 10% ethylene glycol
BCS-F01	25% PEG Smear Low -- 0.1M HEPES pH 7.5 -- 0.15M sodium citrate tribasic
BCS-F02	28% PEG Smear Low -- 0.1M tris pH 8.5 -- 0.2M sodium chloride -- 5% glycerol
BCS-F03	15% PEG Smear Medium -- 0.1M tris pH 8.0 -- 0.075M sodium acetate -- 0.15M sodium chloride
BCS-F04	25% PEG Smear Medium -- 0.1M bis-tris-propane pH 8.5 -- 0.1M sodium chloride -- 0.1M sodium formate
BCS-F05	20% PEG Smear Medium -- 0.1M bicine pH 9.0 -- 0.2M ammonium sulfate -- 0.05M magnesium sulfate
BCS-F06	18% PEG Smear High -- 0.1M bis-tris-propane pH 8.5 -- 0.2M ammonium nitrate
BCS-F07	25% PEG Smear High -- 0.1M tris pH 8.0 -- 0.2M MgCl2 -- 10% glycerol
BCS-F08	28% PEG Smear Broad -- 0.1M tris pH 8.5 -- 0.15M ammonium acetate -- 0.01M CaCl2
BCS-F09	25% PEG Smear Broad -- 0.1M bicine pH 9.0 -- 10% 2-propanol
BCS-F10	20% PEG Smear Broad -- 0.1M tris pH 8.0 -- 0.2M ammonium sulfate
BCS-F11	22.5% PEG Smear Broad -- 0.1M bicine pH 9.0 -- 0.02M magnesium sulfate -- 0.2M KCl
BCS-F12	22.5% PEG Smear Broad -- 0.1M cacodylate pH 5.5 -- 0.1M Na/K tartrate -- 10% ethylene glycol
BCS-G01	14% PEG Smear Low -- 0.1M MES pH 6.0 -- 0.01M cobalt chloride -- 0.1M magnesium formate
BCS-G02	25% PEG Smear Low -- 0.1M bis-tris pH 7.0 -- 0.15M lithium sulfate -- 0.05M MgCl2
BCS-G03	25% PEG Smear Medium -- 0.1M HEPES pH 7.5 -- 0.2M ammonium sulfate -- 0.01M cadmium chloride
BCS-G04	18% PEG Smear Medium -- 0.1M MgCl2 -- 10% ethylene glycol -- 0.1M KCl
BCS-G05	12% PEG Smear Medium -- 0.1M MES pH 6.5 -- 0.1M magnesium acetate -- 10% ethylene glycol
BCS-G06	12% PEG Smear High -- 0.1M MES pH 6.0 -- 0.1M magnesium acetate -- 0.1M KCl
BCS-G07	8% PEG Smear High -- 0.1M PIPES pH 7.0 -- 0.05M CaCl2 -- 0.05M sodium formate
BCS-G08	18% PEG Smear Broad -- 0.1M bis-tris pH 6.0 -- 0.075M MgCl2 -- 0.075M sodium citrate tribasic
BCS-G09	15% PEG Smear Broad -- 0.1M bis-tris pH 6.5 -- 0.1M MgCl2 -- 0.1M sodium acetate
BCS-G10	25% PEG Smear Broad -- 0.1M HEPES pH 7.0 -- 0.1M ammonium sulfate -- 0.1M sodium formate
BCS-G11	22.5% PEG Smear Medium -- 0.1M HEPES pH 7.5 -- 0.2M Na/K phosphate-- 10% glycerol
BCS-G12	22.5% PEG Smear Broad -- 0.1M tris pH 7.5 -- 0.25M sodium chloride -- 0.05M L-Proline
BCS-H01	25% PEG Smear Low -- 0.1M tris pH 8.0 -- 0.05M CaCl2 -- 0.05M sodium formate
BCS-H02	20% PEG Smear Low -- 0.1M PIPES pH 7.0 -- 0.1M MgCl2 -- 0.1M rubidium chloride
BCS-H03	15% PEG Smear Medium -- 0.1M HEPES pH 7.5 -- 0.2M MgCl2 -- 5% 2-propanol -- 10% ethylene glycol
BCS-H04	12% PEG Smear Medium -- 0.1M HEPES pH 7.0 -- 0.05M ammonium acetate -- 0.15M magnesium sulfate
BCS-H05	20% PEG Smear Medium -- 0.1M HEPES pH 7.0 -- 7.5% tacsimate
BCS-H06	15% PEG Smear High -- 0.1M bis-tris pH 7.0 -- 0.1M ammonium acetate -- 0.1M zinc chloride
BCS-H07	20% PEG Smear High -- 0.1M HEPES pH 8.0 -- 0.15M lithium sulfate -- 0.05M MgCl2
BCS-H08	25% PEG Smear Broad -- 0.1M tris pH 8.0 -- 0.1M potassium thiocyanate -- 0.1M sodium bromide
BCS-H09	28% PEG Smear Broad -- 0.1M bis-tris-propane pH 8.5 -- 0.05M ammonium sulfate -- 0.05M lithium sulfate
BCS-H10	15% PEG Smear Broad -- 0.1M PIPES pH 7.0 -- 0.2M ammonium sulfate -- 0.01M CdCl2 -- 10% ethylene glycol
BCS-H11	22.5% PEG Smear Broad -- 0.1M bis-tris pH 7.5 -- 0.2M lithium sulfate -- 0.05M zinc acetate
BCS-H12	22.5% PEG Smear Broad -- 0.1M HEPES pH 8.0 -- 0.075M NaBr -- 0.05M NaF -- 0.075M NaI

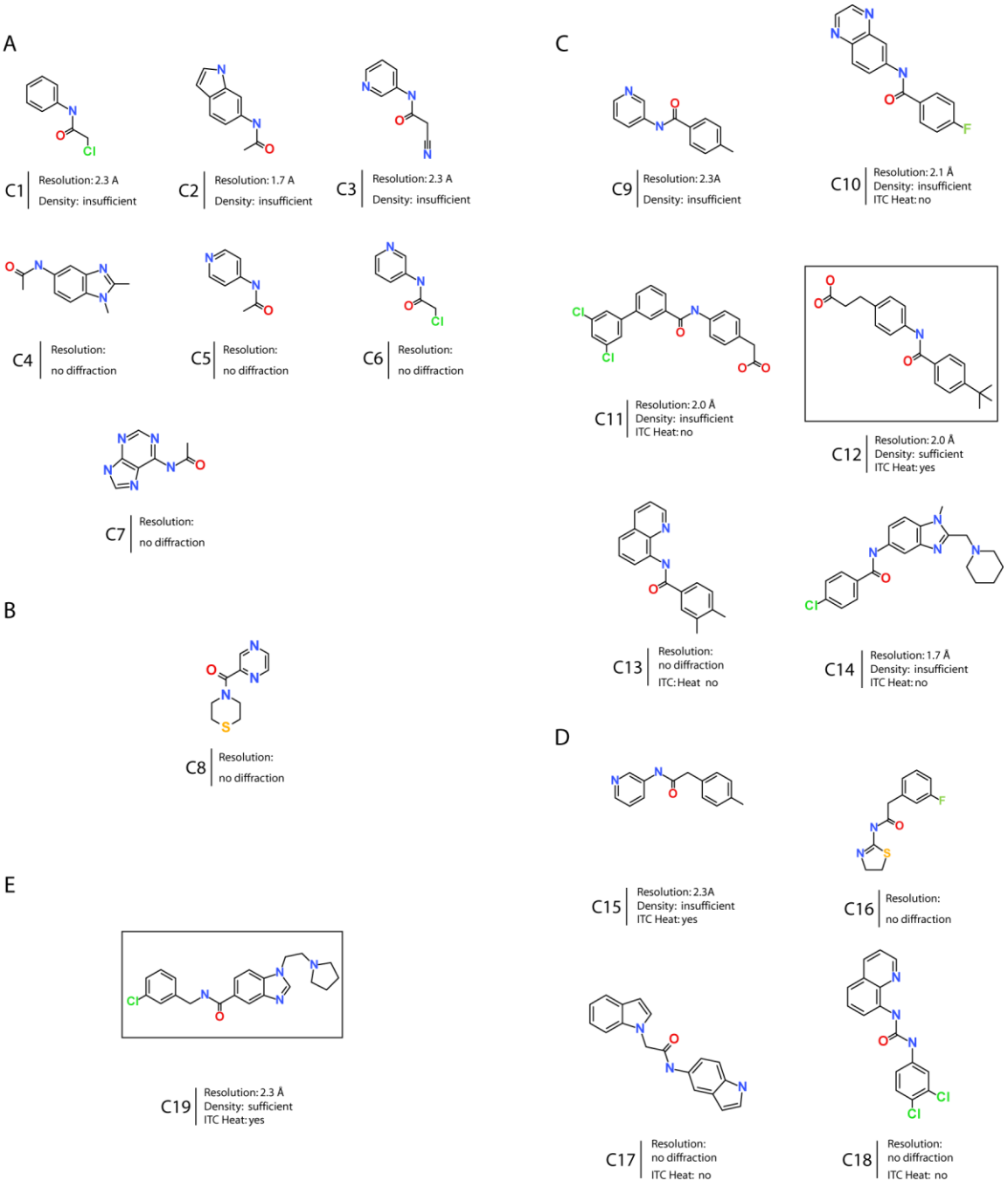
Supplementary Table S6 Composition of the JCSG7 coarse screen

JCSG7-A01	50% PEG400 -- 0.2M lithium sulfate -- 0.1M acetate pH 4.5
JCSG7-A02	20% PEG3000 -- 0.1M citrate pH 5.5
JCSG7-A03	20% PEG3350 -- 0.2M ammonium citrate dibasic
JCSG7-A04	30% MPD -- 0.02M calcium chloride -- 0.1M acetate pH 4.5
JCSG7-A05	20% PEG3350 -- 0.2M magnesium formate
JCSG7-A06	20% PEG1000 -- 0.2M lithium sulfate -- 0.1M citrate pH 4.2
JCSG7-A07	20% PEG8000 -- 0.1M CAPSO pH 9.5
JCSG7-A08	20% PEG3350 -- 0.2M ammonium formate
JCSG7-A09	20% PEG3350 -- 0.2M ammonium chloride
JCSG7-A10	20% PEG3350 -- 0.2M potassium formate
JCSG7-A11	50% MPD -- 0.2M ammonium phosphate dibasic -- 0.1M tris pH 8.5
JCSG7-A12	20% PEG3350 -- 0.2M potassium nitrate
JCSG7-B01	0.8M ammonium sulfate -- 0.1M citrate pH 4.2
JCSG7-B02	20% PEG3350 -- 0.2M potassium thiocyanate
JCSG7-B03	20% PEG6000 -- 0.1M bicine pH 9.0
JCSG7-B04	10% PEG8000 -- 8% ethylene glycol -- 0.1M HEPES pH 7.5
JCSG7-B05	40% MPD -- 5% PEG8000 -- 0.1M cacodylate pH 6.5
JCSG7-B06	5% PEG1000 -- 40% ethanol -- 0.1M citrate pH 4.2
JCSG7-B07	8% PEG4000 -- 0.1M acetate pH 4.5
JCSG7-B08	10% PEG8000 -- 0.2M magnesium chloride -- 0.1M HEPES pH 7.0
JCSG7-B09	20% PEG6000 -- 0.1M citrate pH 5.0
JCSG7-B10	50% PEG200 -- 0.2M magnesium chloride -- 0.1M cacodylate pH 6.5
JCSG7-B11	1.6M sodium citrate tribasic
JCSG7-B12	20% PEG3350 -- 0.2M potassium citrate tribasic
JCSG7-C01	20% PEG8000 -- 0.2M sodium chloride -- 0.1M citrate pH 4.2
JCSG7-C02	20% PEG6000 -- 0.8M lithium chloride -- 0.1M citrate pH 4.2
JCSG7-C03	20% PEG3350 -- 0.2M ammonium nitrate
JCSG7-C04	10% PEG6000 -- 0.1M HEPES pH 7.0
JCSG7-C05	0.8M sodium phosphate monobasic -- 0.8M potassium phosphate dibasic -- 0.1M HEPES pH 7.5
JCSG7-C06	40% PEG300 -- 0.1M citrate pH 4.2
JCSG7-C07	10% PEG3000 -- 0.2M zinc acetate -- 0.1M acetate pH 4.5
JCSG7-C08	20% ethanol -- 0.1M tris pH 8.5
JCSG7-C09	25% 1,2-propanediol -- 10% glycerol -- 0.1M sodium/potassium phosphate pH 7.5
JCSG7-C10	10% PEG20000 -- 2%(v/v) dioxane -- 0.1M bicine pH 9.0
JCSG7-C11	2M ammonium sulfate -- 0.1M acetate pH 4.5
JCSG7-C12	10% PEG1000 -- 10% PEG8000
JCSG7-D01	25% PEG1000 -- 20% glycerol
JCSG7-D02	30% PEG400 -- 0.2M magnesium chloride -- 0.1M HEPES pH 7.5
JCSG7-D03	50% PEG200 -- 0.2M sodium chloride -- 0.1M sodium/potassium phosphate pH 7.5
JCSG7-D04	30% PEG8000 -- 0.2M lithium sulfate -- 0.1M acetate pH 4.5
JCSG7-D05	60% MPD -- 0.1M HEPES pH 7.5
JCSG7-D06	20% PEG8000 -- 0.2M magnesium chloride -- 0.1M tris pH 8.5
JCSG7-D07	40% PEG400 -- 0.2M lithium sulfate -- 0.1M tris pH 8.5
JCSG7-D08	40% MPD -- 0.1M tris pH 8.0
JCSG7-D09	0.15M ammonium sulfate -- 25% PEG4000 -- 15% glycerol
JCSG7-D10	40% PEG300 -- 0.2M calcium acetate -- 0.1M cacodylate pH 6.5
JCSG7-D11	30% glycerol -- 15% 2-propanol -- 0.15M calcium chloride -- 0.1M acetate pH 4.5
JCSG7-D12	16% PEG8000 -- 0.04M potassium phosphate dibasic -- 20% glycerol
JCSG7-E01	1M sodium citrate tribasic -- 0.1M cacodylate pH 6.5
JCSG7-E02	2M ammonium sulfate -- 0.2M sodium chloride -- 0.1M cacodylate pH 6.5
JCSG7-E03	10% 2-propanol -- 0.2M sodium chloride -- 0.1M HEPES pH 7.5
JCSG7-E04	1.26M ammonium sulfate -- 0.2M lithium sulfate -- 0.1M tris pH 8.5
JCSG7-E05	40% MPD -- 0.1M CAPS pH 10.5
JCSG7-E06	20% PEG3000 -- 0.2M zinc acetate -- 0.1M HEPES pH 7.5

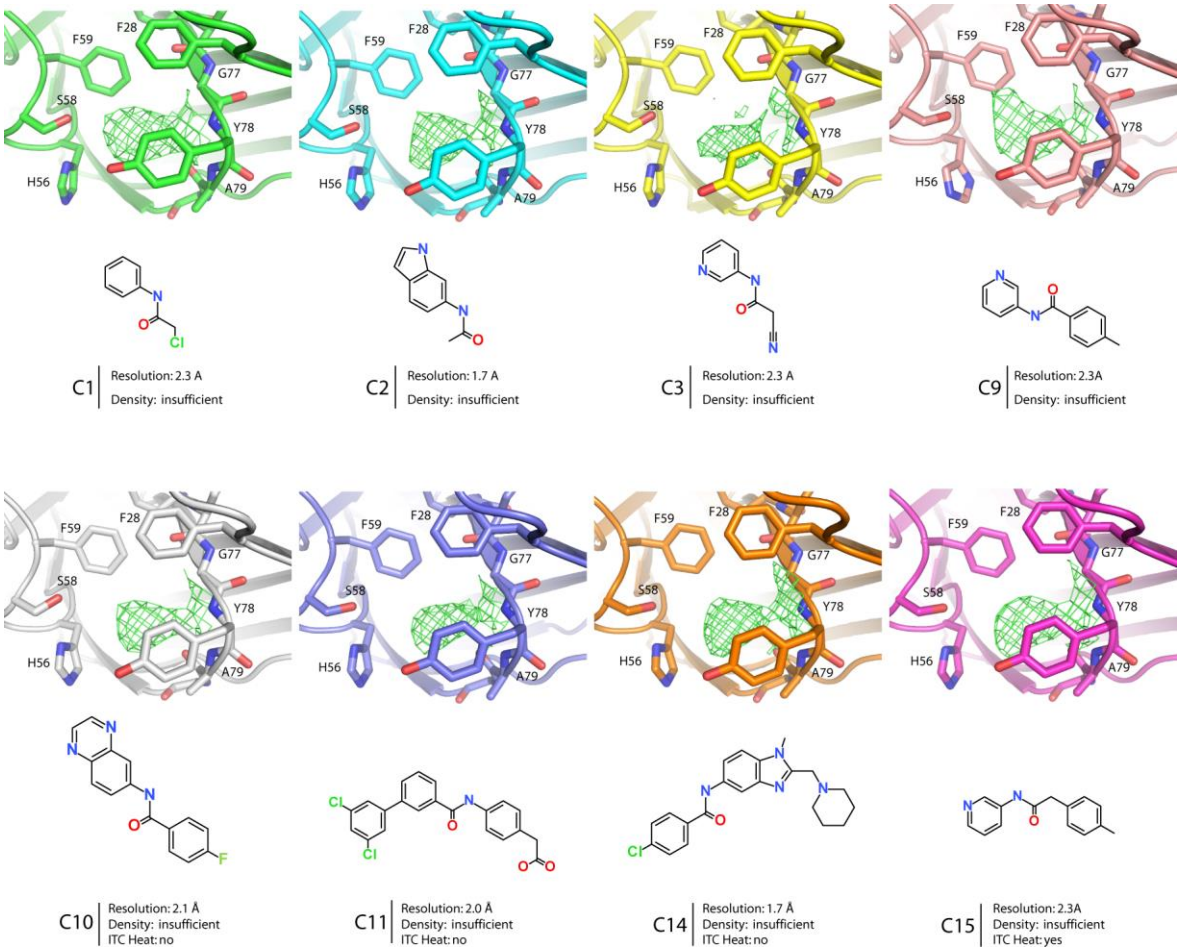
JCSG7-E07	10% 2-propanol -- 0.2M zinc acetate -- 0.1M cacodylate pH 6.5
JCSG7-E08	0.8M ammonium phosphate dibasic -- 0.1M acetate pH 4.5
JCSG7-E09	1.6M magnesium sulfate -- 0.1M MES pH 6.5
JCSG7-E10	10% PEG6000 -- 0.1M bicine pH 9.0
JCSG7-E11	16% PEG8000 -- 20% glycerol -- 0.16M calcium acetate -- 0.1M cacodylate pH 6.5
JCSG7-E12	10% PEG8000 -- 0.1M tris pH 8.0
JCSG7-F01	30% jeffamine M-600 -- 0.05M cesium chloride -- 0.1M MES pH 6.5
JCSG7-F02	3M ammonium sulfate -- 0.1M citrate pH 5.0
JCSG7-F03	20% MPD -- 0.1M tris pH 8.0
JCSG7-F04	20% jeffamine M-600 -- 0.1M HEPES pH 7.5
JCSG7-F05	50% ethylene glycol -- 0.2M magnesium chloride -- 0.1M tris pH 8.5
JCSG7-F06	10% MPD -- 0.1M bicine pH 9.0
JCSG7-F07	0.8M succinic acid
JCSG7-F08	2.1M DL- malic acid
JCSG7-F09	2.4M sodium malonate
JCSG7-F10	1.2M sodium malonate -- 0.5% jeffamine ED-2003 -- 0.1M HEPES pH 7.0
JCSG7-F11	1M succinic acid -- 1% PEG2000MME -- 0.1M HEPES pH 7.0
JCSG7-F12	30% jeffamine M-600 -- 0.1M HEPES pH 7.0
JCSG7-G01	30% jeffamine ED-2003 -- 0.1M HEPES pH 7.0
JCSG7-G02	22% polyacrylic acid 5100 -- 0.02M magnesium chloride -- 0.1M HEPES pH 7.5
JCSG7-G03	20% polyvinylpyrrolidone -- 0.01M cobalt chloride -- 0.1M tris pH 8.5
JCSG7-G04	20% PEG2000MME -- 0.2M trimethylamine N-oxide -- 0.1M tris pH 8.5
JCSG7-G05	12% PEG3350 -- 0.005M (each) Co(II)Cl2, CdCl, Ni(II)Cl2, MgCl2 -- 0.1M HEPES pH 7.5
JCSG7-G06	20% PEG3350 -- 0.2M sodium malonate
JCSG7-G07	20% PEG3350 -- 0.1M succinic acid
JCSG7-G08	20% PEG3350 -- 0.15M DL- malic acid
JCSG7-G09	30% PEG2000MME -- 0.1M potassium thiocyanate
JCSG7-G10	30% PEG2000MME -- 0.15M potassium bromide
JCSG7-G11	2M ammonium sulfate -- 0.1M bis-tris pH 5.5
JCSG7-G12	3M sodium chloride -- 0.1M bis-tris pH 5.5
JCSG7-H01	0.3M magnesium formate -- 0.1M bis-tris pH 5.5
JCSG7-H02	1% PEG3350 -- 1M ammonium sulfate -- 0.1M bis-tris pH 5.5
JCSG7-H03	25% PEG3350 -- 0.1M bis-tris pH 5.5
JCSG7-H04	45% MPD -- 0.2M calcium chloride -- 0.1M bis-tris pH 5.5
JCSG7-H05	45% MPD -- 0.2M ammonium acetate -- 0.1M bis-tris pH 5.5
JCSG7-H06	0.1M ammonium acetate -- 0.1M bis-tris pH 5.5 -- 16%(w/v) PEG10000
JCSG7-H07	25% PEG3350 -- 0.2M ammonium sulfate -- 0.1M bis-tris pH 5.5
JCSG7-H08	25% PEG3350 -- 0.2M sodium chloride -- 0.1M bis-tris pH 5.5
JCSG7-H09	25% PEG3350 -- 0.2M lithium sulfate -- 0.1M bis-tris pH 5.5
JCSG7-H10	25% PEG3350 -- 0.2M ammonium acetate -- 0.1M bis-tris pH 5.5
JCSG7-H11	25% PEG3350 -- 0.2M magnesium chloride -- 0.1M bis-tris pH 5.5
JCSG7-H12	45% MPD -- 0.2M ammonium acetate -- 0.1M HEPES pH 7.5



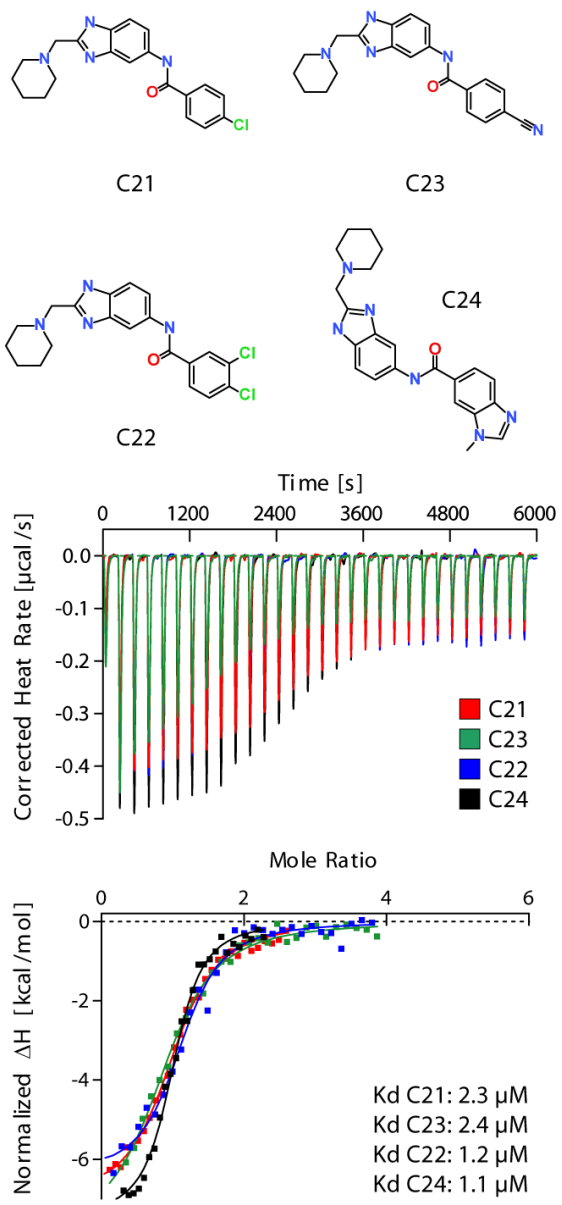
Supplementary Figure S3 Overlayed binding modes of JQ1 (green) in BRD4(1) (PDB ID: 3MXF) and PFI-3 (grey) in SMARCA2 (PDB ID: 5DKC) (*6, 199*). Surface representation of the BRD4(1) binding pocket exemplifies the shape of a BRD binding pocket. Water molecules are shown as red spheres in the JQ1-bound structure and as cyan sphere in the PFI-3-bound structure. PFI-3 protrudes deeper into the BRD binding pocket as JQ1 and depletes four conserved water molecules.



Supplementary Figure S4 Structures of the selected 19 fragment-like compounds used in this study. Compounds were tested their binding using protein crystallography and ITC, of which the results are indicated. (compound C12 synthesised by J. Schmidt, Figure adapted from (*22*))



Supplementary Figure S5 Analyses of electron density maps within the binding pocket of ENL complexed with the studied ligands. In comparison to the apo-structure, additional density in proximity to the binding site of the acetyl-lysine is observed in all complexes, suggesting the presence of the Kac mimetic amide group (adapted from (*22*))



Supplementary Figure S6 ITC data for the interactions between ENL and compounds C21-C24. Shown are the structures of initial hits (top) panel as well as isothermal titration calorimetry (ITC) data (lower panel). The ITC data are depicted as raw binding heats for compounds C21-C24 coloured as indicated in the figure, as well as normalised binding enthalpies and fitted binding isotherms (single binding site model). The dissociation constants K_D are given in the Figure (compounds synthesised by M. Moustakim, Figure adapted from (*22*))

Supplementary Table S7 Compound SMILES

Compound	SMILE
A1	CC1=C2C(OC(=O)C3=C2C=CC=C3)=NN1
A2	CCC1=C2C(OC(=O)C3=C2C=CC=C3)=NN1
A3	CCCC1=C2C(OC(=O)C3=C2C=CC=C3)=NN1
A4	CCCCC1=C2C(OC(=O)C3=C2C=CC=C3)=NN1
JQ1	CC1=C(SC2=C1C(=NC(C3=NN=C(N32)C)CC(=O)OC(C)(C)C)C4=CC=C(C=C4)Cl)C
PFI-3	C1C2CN(C1CN2C3=CC=CC=N3)C=CC(=O)C4=CC=CC=C4O
BI-2536	CCC1C(=O)N(C2=CN=C(N=C2N1C3CCCC3)NC4=C(C=C(C=C4)C(=O)NC5CCN(CC5)C)OC)C
B1	O=C1[C@@H](CC)N(CC2=CC=CC=C2)C3=C(N1C)C=NC(NC4=C(OCC)C=C(C5CCN(C)CC5)C=C4)=N3
B2	O=C1[C@@H](CC)N(CC2=C(C)C=CS2)C3=C(N1C)C=NC(NC4=C(OCC)C=C(C5CCN(C)CC5)C=C4)=N3
B3	O=C1[C@@H](CC)N(CC2=CC(C)=CS2)C3=C(N1C)C=NC(NC4=C(OCC)C=C(C5CCN(C)CC5)C=C4)=N3
C1	ClCC(=O)Nc1ccccc1
C2	CC(=O)Nc1ccc2cc[nH]c2c1
C3	O=C(CC#N)Nc1cccnc1
C4	CC1N=C2C=C(NC(=O)C)C=CC2N1C
C5	CC(=O)Nc1ccncc1
C6	ClCC(=O)Nc1cccnc1
C7	CC(=O)Nc1ncnc2[nH]cnc12
C8	O=C(N1CCSCC1)c2cnccn2
C9	Cc1ccc(cc1)C(=O)Nc2cccnc2
C10	Fc1ccc(cc1)C(=O)Nc2ccc3nccnc3c2
C11	OC(=O)Cc1ccc(NC(=O)c2cccc(c2)c3cc(Cl)cc(Cl)c3)cc1
C12	CC(C)(C)c1ccc(cc1)C(=O)Nc2ccc(CCC(=O)O)cc2
C13	Cc1ccc(cc1C)C(=O)Nc2cccc3cccnc23
C14	Cn1c(CN2CCCCC2)nc3cc(NC(=O)c4ccc(Cl)cc4)ccc13
C15	Cc1ccc(CC(=O)Nc2cccnc2)cc1
C16	Fc1cccc(CC(=O)NC2=NCCS2)c1
C17	O=C(Cn1ccc2ccccc12)Nc3ccc4[nH]ccc4c3
C18	Clc1ccc(NC(=O)Nc2cccc3cccnc23)cc1Cl
C19	Clc1cccc(CNC(=O)c2ccc3c(c2)ncn3CCN4CCCC4)c1
C20	Cc1ccc(cc1I)C(=O)Nc2ccc3[nH]c(CN4CCCCC4)nc3c2
C21	Clc1ccc(cc1)C(=O)Nc1ccc2[nH]c(CN3CCCCC3)nc2c1
C22	Clc1ccc(cc1Cl)C(=O)Nc1ccc2[nH]c(CN3CCCCC3)nc2c1
C23	O=C(Nc1ccc2[nH]c(CN3CCCCC3)nc2c1)c1ccc(cc1)C#N
C24	Cn1cnc2ccc(cc12)C(=O)Nc1ccc2[nH]c(CN3CCCCC3)nc2c1
SGC-iMLLT	CC1CCCN1CC2=NC3=C(N2)C=C(C=C3)NC(=O)C4=CC5=C(C=C4)N(N=C5)C
C94	CC1CCCN1CC2=NC3=C([N]2)C=C(C=C3)NC(=O)C4=CC5=C(C=C4)[N](N=C5)C6CC6
C25	C(C1=CN=C(C=C1)OC)NC(=O)N2CCN(CC2)C3=CC=C(C=C3)C(C)=O
C26	C(C1=CC=C(C=C1)CN2CCCC2)NC(=O)N3CCN(CC3)C(C4=CC=CS4)=O

Abbreviations

5Aza	5-aza-2'-deoxycytidine
abl	The Abelson proto-oncogene
AF9	ALL1-fused gene from chromosome 9 protein
AFF4	AF4/FMR2 Family Member 4
ALK	Anaplastic lymphoma kinase
AML	Acute myeloid leukemia
APS	Ammonium persulfate
ATAC	ADA2A-containing complex
ATAD2	ATPase family AAA domain-containing protein 2
ATP	Adenosine triphosphate
BAF	Brg/Brahma-associated factor
BAZ	Bromodomain adjacent to Zinc Finger Domain
BET	Bromo- and extra-terminal
Bhb	β-hydroxybutyrylation
BRD	Bromodomain
BRDT	Bromodomain testis-specific protein
BRET	Bioluminescence resonance energy transfer
BRPF1-3	Bromodomain and PHD Finger containing 1-3
BZD	Benzodiazepine
CBP	CREB-binding protein
CCD	Charged-coupled devices
CDK9	Cyclin-dependent kinase 9
CECR2	Cat eye syndrome critical region protein 2
c-MET	Mesenchymal-epithelial transition factor
CMOS	Complementary metal-oxide-semiconductor
CoA	Coenzyme A
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTIP2	COUP-TF-interacting protein 2
CV	Column volume
DANN	Deoxyribonucleic acid
dCMP	Deoxycytidine monophosphate
dCTP	Deoxycytidine triphosphate
dGMP	Deoxyguanosine monophosphate
dGTP	Deoxyguanosine triphosphate
DMEM	Dulbecco's modified Eagle's medium
DNMT	DNA methyltransferase
dNTP	Deoxynucleoside triphosphate
DOT1L	Disruptor of telomeric silencing 1-like
DSF	Differential scanning fluorimetry
EBAF	ENL-associated BAF-containing BAF250b
ENL	Eleven-nineteen-leukemia protein
ESI-TOF	Electrospray ionization time of flight
ETS1	Protein C-ets-1

FALZ	Fetal Alz-50 clone 1 protein
FOM	Figure of merit
FRAP	Fluorescence recovery after photobleaching
GAS41	Glioma-amplified sequence 41
GCN5	General Control Nonderepressible 5
GFP	Green fluorescent protein
GLUT1	Glucose transporter I
GUI	Graphical user interface
HAT	Histone acetyl transferase
HDAC	Histone deacetylase
HEXIM1	Hexamethylene bis-acetamide-inducible protein 1
His_6 tag	Hexahistidine tag
HTS	High-throughput screening
ICR	Institute of Cancer Research
IPTG	Isopropyl-β-D-thiogalactopyranosid
ITC	Isothermal titration calorimetry
LB	Lysogeny broth
LBDD	Ligand-based drug discovery
LIC	Ligation-independent cloning
LSD1	Lysine-specific histone demethylase 1A
LSD1	Spindlin1 and Lysine-specific histone demethylase 1A
MAP2K1	MAP kinase kinase 1
MBT	Malignant brain tumour
MLL	Mixed-lineage leukemia
MLLT1/3	Myeloid/lymphoid or mixed-lineage leukemia translocated to chromosome 1/3
MPN	Myeloproliferative neoplasms
mRNA	Messenger ribonucleic acid
MWCO	Molecular weight cutoff
MYST	MOZ, Ybf2/Sas3, SAS2 and Tip60
NCI/DTP	National Cancer Institute/Developmental Therapeutics Program
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear magnetic resonance
NSCLC	Non-small cell lung cancer
OD	Optical density
PAF1	RNA polymerase II-associated factor 1 homolog
PB1	Polybromo-1
PBAF	Polybromo-associated BRG1 factor
PCAF	P300/CBP-associated factor
PCR	Polymerase chain reaction
PDB	Protein data bank
PDGF-R	Platelet-derived growth factor receptor
PEG	Polyethylene glycol
PHD	Plant homeodomain
PI3K	Phosphoinositide 3-kinase
PKMT	Protein lysine methyltransferase

PML	Promyelocytic leukemia
Pol II	RNA polymerase II
PRMT	Protein arginine methyltransferase
P-TEFb	Positive elongation factor b
PTM	Post-translational modification
PWWP	pro-trp-trp-pro
QSAR	Quantitative structure-activity relationship
ROS1	c-ros oncogene 1
rpm	rpm
SAHA	Suberoylanilide hydroxamic acid
SAR	Structure-activity relationships
SBDD	Structure-based drug discovery
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SET	Su(var)3-9, Enhancer-of-zeste and Trithorax
SGC	Structural Genomics Consortium
SIRT1	Silent mating type information regulation-1
SMARCA2/4	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator of Chromatin, Subfamily A, Member 2/4
SMYD2	SET and MYND domain-containing protein 2
SPR	Surface plasmon resonance
SRCAP	Snf2-related CBP activator protein remodelling complex
ss	Single-stranded
SWI/SNF	Switch/Sucrose Non-Fermentable
TAF1	Transcription initiation factor TFIID subunit 1
TB	Terrific broth
TCEP	Tris(2−carboxyethyl)phosphine
TEMED	Tetramethylethylenediamine
TEV	Tobacco etch virus
TNBC	Triple-negative breast cancer
TRIM	Tripartite motif-containing protein
YEATS	Yaf9, ENL, AF9, Taf14, Sas5
YEATS2/4	YEATS domain-containing protein 2/4

References

1. Chen, Z.; Li, S.; Subramaniam, S.; Shyy, J. Y.-J.; Chien, S. Epigenetic Regulation: A New Frontier for Biomedical Engineers. *Annual review of biomedical engineering [Online]* **2017,** *19*, 195–219.
2. Ferrell, J. E. Bistability, bifurcations, and Waddington's epigenetic landscape. *Current biology : CB [Online]* **2012,** *22* (11), R458-66.
3. Portela, A.; Esteller, M. Epigenetic modifications and human disease. *Nature Biotechnology [Online]* **2010,** *28* (10), 1057. https://www.nature.com/articles/nbt.1685.pdf.
4. Shahbazian, M. D.; Grunstein, M. Functions of site-specific histone acetylation and deacetylation. *Annual review of biochemistry [Online]* **2007,** *76*, 75–100.
5. Owen, D. J.; Ornaghi, P.; Yang, J.-C.; Lowe, N.; Evans, P. R.; Ballario, P.; Neuhaus, D.; Filetici, P.; Travers, A. A. The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase Gcn5p. *The EMBO Journal [Online]* **2000,** *19* (22), 6141–6149.
6. Filippakopoulos, P.; Picaud, S.; Mangos, M.; Keates, T.; Lambert, J.-P.; Barsyte-Lovejoy, D.; Felletar, I.; Volkmer, R.; Müller, S.; Pawson, T.; Gingras, A.-C.; Arrowsmith, C. H.; Knapp, S. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell [Online]* **2012,** *149* (1), 214–231.
7. Li, Y.; Wen, H.; Xi, Y.; Tanaka, K.; Wang, H.; Peng, D.; Ren, Y.; Jin, Q.; Dent, S. Y. R.; Li, W.; Li, H.; Shi, X. AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. *Cell [Online]* **2014,** *159* (3), 558–571.
8. Muller, S.; Filippakopoulos, P.; Knapp, S. Bromodomains as therapeutic targets. *Expert reviews in molecular medicine [Online]* **2011,** *13*, e29.
9. Filippakopoulos, P.; Qi, J.; Picaud, S.; Shen, Y.; Smith, W. B.; Fedorov, O.; Morse, E. M.; Keates, T.; Hickman, T. T.; Felletar, I.; Philpott, M.; Munro, S.; McKeown, M. R.; Wang, Y.; Christie, A. L.; West, N.; Cameron, M. J.; Schwartz, B.; Heightman, T. D.; La Thangue, N.; French, C. A.; Wiest, O.; Kung, A. L.; Knapp, S.; Bradner, J. E. Selective inhibition of BET bromodomains. *Nature [Online]* **2010,** *468* (7327), 1067–1073.
10. Zhao, D.; Li, Y.; Xiong, X.; Chen, Z.; Li, H. YEATS Domain-A Histone Acylation Reader in Health and Disease. *Journal of molecular biology [Online]* **2017,** *429* (13), 1994–2002.
11. 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.; Buser-Doepner, C.; Campbell, R. M.; Carter, A. J.; Cohen, P.; Copeland, R. A.; Cravatt, B.; Dahlin, J. L.; Dhanak, D.; Edwards, A. M.; Frederiksen, M.; Frye, S. V.; Gray, N.; Grimshaw, C. E.; Hepworth, D.; Howe, T.; Huber, K. V. M.; Jin, J.; Knapp, S.; Kotz, J. D.; Kruger, R. G.; Lowe, D.; Mader, M. M.; Marsden, B.; Mueller-Fahrnow, A.; Müller, S.; O'Hagan, R. C.; Overington, J. P.; Owen, D. R.; Rosenberg, S. H.; Roth, B.; Roth, B.; Ross, R.; Schapira, M.; Schreiber, S. L.; Shoichet, B.; Sundström, M.; Superti-Furga, G.; Taunton, J.; Toledo-Sherman, L.; Walpole, C.; Walters, M. A.; Willson, T. M.; Workman, P.; Young, R. N.; Zuercher, W. J. The promise and peril of chemical probes. *Nature chemical biology [Online]* **2015,** *11* (8), 536–541.

12. Wu, Q.; Heidenreich, D.; Zhou, S.; Ackloo, S.; Krämer, A.; Nakka, K.; Lima-Fernandes, E.; Deblois, G.; Duan, S.; Dilworth, J.; Lupien, M.; Brennan, P. E.; Arrowsmith, C. H.; Mueller, S.; Fedorov, O.; Filippakopoulos, P.; Knapp, S. A chemical toolbox for the study of bromodomains and epigenetic signaling. *Nature Communications [Online]* **2019**, 10(1), 1915.

13. Myrianthopoulos, V.; Gaboriaud-Kolar, N.; Tallant, C.; Hall, M.-L.; Grigoriou, S.; Brownlee, P. M.; Fedorov, O.; Rogers, C.; Heidenreich, D.; Wanior, M.; Drosos, N.; Mexia, N.; Savitsky, P.; Bagratuni, T.; Kastritis, E.; Terpos, E.; Filippakopoulos, P.; Müller, S.; Skaltsounis, A.-L.; Downs, J. A.; Knapp, S.; Mikros, E. Discovery and Optimization of a Selective Ligand for the Switch/Sucrose Nonfermenting-Related Bromodomains of Polybromo Protein-1 by the Use of Virtual Screening and Hydration Analysis. *Journal of medicinal chemistry [Online]* **2016,** *59* (19), 8787–8803.

14. Ciceri, P.; Müller, S.; O'Mahony, A.; Fedorov, O.; Filippakopoulos, P.; Hunt, J. P.; Lasater, E. A.; Pallares, G.; Picaud, S.; Wells, C.; Martin, S.; Wodicka, L. M.; Shah, N. P.; Treiber, D. K.; Knapp, S. Dual kinase-bromodomain inhibitors for rationally designed polypharmacology. *Nature chemical biology [Online]* **2014,** *10* (4), 305. https://www.nature.com/articles/nchembio.1471.pdf.

15. Ferri, E.; Petosa, C.; McKenna, C. E. Bromodomains: Structure, function and pharmacology of inhibition. *Biochemical pharmacology [Online]* **2016,** *106*, 1–18.

16. Watts, E.; Heidenreich, D.; Tucker, E.; Raab, M.; Strebhardt, K.; Chesler, L.; Knapp, S.; Bellenie, B.; Hoelder, S. Designing Dual Inhibitors of Anaplastic Lymphoma Kinase (ALK) and Bromodomain-4 (BRD4) by Tuning Kinase Selectivity. *Journal of medicinal chemistry [Online]* **2019,** *62* (5), 2618–2637.

17. Li, Y.; Sabari, B. R.; Panchenko, T.; Wen, H.; Zhao, D.; Guan, H.; Wan, L.; Huang, H.; Tang, Z.; Zhao, Y.; Roeder, R. G.; Shi, X.; Allis, C. D.; Li, H. Molecular Coupling of Histone Crotonylation and Active Transcription by AF9 YEATS Domain. *Molecular Cell [Online]* **2016,** *62* (2), 181–193.

18. Zhao, D.; Guan, H.; Zhao, S.; Mi, W.; Wen, H.; Li, Y.; Zhao, Y.; Allis, C. D.; Shi, X.; Li, H. YEATS2 is a selective histone crotonylation reader. *Cell research [Online]* **2016,** *26* (5), 629–632.

19. Wan, L.; Wen, H.; Li, Y.; Lyu, J.; Xi, Y.; Hoshii, T.; Joseph, J. K.; Wang, X.; Loh, Y.-H. E.; Erb, M. A.; Souza, A. L.; Bradner, J. E.; Shen, L.; Li, W.; Li, H.; Allis, C. D.; Armstrong, S. A.; Shi, X. ENL links histone acetylation to oncogenic gene expression in acute myeloid leukaemia. *Nature [Online]* **2017,** *543* (7644), 265–269.

20. Wang, Y.; Jin, J.; Chung, M. W. H.; Feng, L.; Sun, H.; Hao, Q. Identification of the YEATS domain of GAS41 as a pH-dependent reader of histone succinylation. *Proceedings of the National Academy of Sciences of the United States of America [Online]* **2018,** *115* (10), 2365–2370.

21. Erb, M. A.; Scott, T. G.; Li, B. E.; Xie, H.; Paulk, J.; Seo, H.-S.; Souza, A.; Roberts, J. M.; Dastjerdi, S.; Buckley, D. L.; Sanjana, N. E.; Shalem, O.; Nabet, B.; Zeid, R.; Offei-Addo, N. K.; Dhe-Paganon, S.; Zhang, F.; Orkin, S. H.; Winter, G. E.; Bradner, J. E. Transcription control by the ENL YEATS domain in acute leukaemia. *Nature [Online]* **2017,** *543* (7644), 270–274.

22. Heidenreich, D.; Moustakim, M.; Schmidt, J.; Merk, D.; Brennan, P. E.; Fedorov, O.; Chaikuad, A.; Knapp, S. Structure-Based Approach toward Identification of Inhibitory Fragments for Eleven-Nineteen-Leukemia Protein (ENL). *Journal of medicinal chemistry [Online]* **2018**, 61 (23), 10929-10934

23. Christott, T.; Bennett, J.; Coxon, C.; Monteiro, O.; Giroud, C.; Beke, V.; Felce, S. L.; Gamble, V.; Gileadi, C.; Poda, G.; Al-Awar, R.; Farnie, G.; Fedorov, O. Discovery of a Selective Inhibitor for the YEATS Domains of ENL/AF9. *SLAS discovery : advancing life sciences R & D [Online]* **2018**, 2472555218809904.

24. Moustakim, M.; Christott, T.; Monteiro, O. P.; Bennett, J.; Giroud, C.; Ward, J.; Rogers, C. M.; Smith, P.; Panagakou, I.; Díaz-Sáez, L.; Felce, S. L.; Gamble, V.; Gileadi, C.; Halidi, N.; Heidenreich, D.; Chaikuad, A.; Knapp, S.; Huber, K. V. M.; Farnie, G.; Heer, J.; Manevski, N.; Poda, G.; Al-Awar, R.; Dixon, D. J.; Brennan, P. E.; Fedorov, O. Discovery of an MLLT1/3 YEATS Domain Chemical Probe. *Angewandte Chemie (International ed. in English) [Online]* **2018,** *57* (50), 16302–16307.

25. Pervaiz, M.; Mishra, P.; Günther, S. Bromodomain Drug Discovery - the Past, the Present, and the Future. *Chemical record (New York, N.Y.) [Online]* **2018,** *18* (12), 1808–1817.

26. Vidler, L. R.; Brown, N.; Knapp, S.; Hoelder, S. Druggability analysis and structural classification of bromodomain acetyl-lysine binding sites. *Journal of medicinal chemistry [Online]* **2012,** *55* (17), 7346–7359.

27. Watson, J. D.; Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature [Online]* **1953,** *171* (4356), 737–738.

28. Crick, F. Central dogma of molecular biology. *Nature [Online]* **1970,** *227* (5258), 561–563.

29. Wilson, B. J.; Nicholls, S. G. The Human Genome Project, and recent advances in personalized genomics. *Risk management and healthcare policy [Online]* **2015,** *8*, 9–20.

30. Besser, J.; Carleton, H. A.; Gerner-Smidt, P.; Lindsey, R. L.; Trees, E. Next-Generation Sequencing Technologies and their Application to the Study and Control of Bacterial Infections. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases [Online]* **2017,** *24* (4), 335–341.

31. Stoeger, T.; Gerlach, M.; Morimoto, R. I.; Nunes Amaral, L. A. Large-scale investigation of the reasons why potentially important genes are ignored. *PLoS biology [Online]* **2018,** *16* (9), e2006643.

32. Djebali, S.; Davis, C. A.; Merkel, A.; Dobin, A.; Lassmann, T.; Mortazavi, A.; Tanzer, A.; Lagarde, J.; Lin, W.; Schlesinger, F.; Xue, C.; Marinov, G. K.; Khatun, J.; Williams, B. A.; Zaleski, C.; Rozowsky, J.; Röder, M.; Kokocinski, F.; Abdelhamid, R. F.; Alioto, T.; Antoshechkin, I.; Baer, M. T.; Bar, N. S.; Batut, P.; Bell, K.; Bell, I.; Chakrabortty, S.; Chen, X.; Chrast, J.; Curado, J.; Derrien, T.; Drenkow, J.; Dumais, E.; Dumais, J.; Duttagupta, R.; Falconnet, E.; Fastuca, M.; Fejes-Toth, K.; Ferreira, P.; Foissac, S.; Fullwood, M. J.; Gao, H.; Gonzalez, D.; Gordon, A.; Gunawardena, H.; Howald, C.; Jha, S.; Johnson, R.; Kapranov, P.; King, B.; Kingswood, C.; Luo, O. J.; Park, E.; Persaud, K.; Preall, J. B.; Ribeca, P.; Risk, B.; Robyr, D.; Sammeth, M.; Schaffer, L.; See, L.-H.; Shahab, A.; Skancke, J.; Suzuki, A. M.; Takahashi, H.; Tilgner, H.; Trout, D.; Walters, N.; Wang, H.; Wrobel, J.; Yu, Y.; Ruan, X.; Hayashizaki, Y.; Harrow, J.; Gerstein, M.; Hubbard, T.; Reymond, A.; Antonarakis, S. E.; Hannon, G.; Giddings, M. C.; Ruan, Y.; Wold, B.; Carninci, P.; Guigó, R.; Gingeras, T. R. Landscape of transcription in human cells. *Nature [Online]* **2012,** *489* (7414), 101. https://www.nature.com/articles/nature11233.pdf.

33. Weinhold, B. Epigenetics: The Science of Change. *Environmental Health Perspectives [Online]* **2006,** *114* (3), A160-7.

34. Morey, C.; Avner, P. Genetics and epigenetics of the X chromosome. *Annals of the New York Academy of Sciences [Online]* **2010,** *1214*, E18-33.

35. Gileadi, O.; Knapp, S.; Lee, W. H.; Marsden, B. D.; Müller, S.; Niesen, F. H.; Kavanagh, K. L.; Ball, L. J.; Delft, F. von; Doyle, D. A.; Oppermann, U. C. T.; Sundström, M. The scientific impact of the Structural Genomics Consortium: a protein family and ligand-centered approach to medically-relevant human proteins. *Journal of structural and functional genomics [Online]* **2007,** *8* (2-3), 107–119.

36. Waddington, C. H. The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser. *The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser. [Online]* **1957**.

37. Turner, B. M. Epigenetic responses to environmental change and their evolutionary implications. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences [Online]* **2009,** *364* (1534), 3403–3418.

38. Schulze, J. M.; Wang, A. Y.; Kobor, M. S. YEATS domain proteins: a diverse family with many links to chromatin modification and transcription. *Biochemistry and cell biology = Biochimie et biologie cellulaire [Online]* **2009,** *87* (1), 65–75.

39. Dekker, F. J.; van den Bosch, T.; Martin, N. I. Small molecule inhibitors of histone acetyltransferases and deacetylases are potential drugs for inflammatory diseases. *Drug discovery today [Online]* **2014,** *19* (5), 654–660.

40. Eckschlager, T.; Plch, J.; Stiborova, M.; Hrabeta, J. Histone Deacetylase Inhibitors as Anticancer Drugs. *International journal of molecular sciences* **2017,** *18* (7). DOI: 10.3390/ijms18071414.

41. Scheer, S.; Ackloo, S.; Medina, T. S.; Schapira, M.; Li, F.; Ward, J. A.; Lewis, A. M.; Northrop, J. P.; Richardson, P. L.; Kaniskan, H. Ü.; Shen, Y.; Liu, J.; Smil, D.; McLeod, D.; Zepeda-Velazquez, C. A.; Luo, M.; Jin, J.; Barsyte-Lovejoy, D.; Huber, K. V. M.; Carvalho, D. D. D.; Vedadi, M.; Zaph, C.; Brown, P. J.; Arrowsmith, C. H. A chemical biology toolbox to study protein methyltransferases and epigenetic signaling. *Nature Communications [Online]***,** *10* (1), 19. https://www.nature.com/articles/s41467-018-07905-4.pdf.

42. Lu, R.; Wang, G. G. Tudor: a versatile family of histone methylation 'readers'. *Trends in biochemical sciences* **2013,** *38* (11). DOI: 10.1016/j.tibs.2013.08.002.

43. Xi, Q.; Wang, Z.; Zaromytidou, A.-I.; Zhang, X. H.-F.; Chow-Tsang, L.-F.; Liu, J. X.; Kim, H.; Barlas, A.; Manova-Todorova, K.; Kaartinen, V.; Studer, L.; Mark, W.; Patel, D. J.; Massagué, J. A poised chromatin platform for TGF-β access to master regulators. *Cell [Online]* **2011,** *147* (7), 1511–1524.

44. Structural Genomics Consortium (SGC). Epigenetics Probe Collection. https://www.thesgc.org/chemical-probes/epigenetics (accessed January 22, 2019).

45. Dutta, A.; Abmayr, S. M.; Workman, J. L. Diverse Activities of Histone Acylations Connect Metabolism to Chromatin Function. *Molecular Cell [Online]* **2016,** *63* (4), 547–552. http://www.sciencedirect.com/science/article/pii/S1097276516303264.

46. Xie, Z.; Di Zhang; Chung, D.; Tang, Z.; Huang, H.; Dai, L.; Qi, S.; Li, J.; Colak, G.; Chen, Y.; Xia, C.; Peng, C.; Ruan, H.; Kirkey, M.; Wang, D.; Jensen, L. M.; Kwon, O. K.; Lee, S.; Pletcher, S. D.; Tan, M.; Lombard, D. B.; White, K. P.; Zhao, H.; Li, J.; Roeder, R. G.; Yang, X.; Zhao, Y. Metabolic Regulation of

Gene Expression by Histone Lysine β-Hydroxybutyrylation. *Molecular Cell [Online]* **2016,** *62* (2), 194–206.

47. Cahill, G. F. Fuel metabolism in starvation. *Annual review of nutrition [Online]* **2006,** *26*, 1–22.
48. Blagg, J.; Workman, P. Choose and Use Your Chemical Probe Wisely to Explore Cancer Biology. *Cancer cell [Online]* **2017,** *32* (1), 9–25.
49. Haque, S. M.; Ratemi, E. S. Drug Development and Analysis Review. *Pharmaceutical Chemistry Journal [Online]* **2017,** *50* (12), 837–850. https://doi.org/10.1007/s11094-017-1543-1.
50. Boettcher, M.; McManus, M. T. Choosing the Right Tool for the Job: RNAi, TALEN or CRISPR. *Molecular Cell [Online]* **2015,** *58* (4), 575–585.
51. Bunnage, M. E.; Chekler, E. L. P.; Jones, L. H. Target validation using chemical probes. *Nature chemical biology [Online]* **2013,** *9* (4), 195–199.
52. Bianchi, V.; Gherardini, P. F.; Helmer-Citterich, M.; Ausiello, G. Identification of binding pockets in protein structures using a knowledge-based potential derived from local structural similarities. *BMC Bioinformatics [Online]* **2012,** *13* (Suppl 4), S17.
53. Acharya, C.; Coop, A.; Polli, J. E.; MacKerell, A. D. Recent Advances in Ligand-Based Drug Design: Relevance and Utility of the Conformationally Sampled Pharmacophore Approach. *Current computer-aided drug design [Online]* **2011,** *7* (1), 10–22.
54. van Montfort, R. L.M.; Workman, P. Structure-based drug design: aiming for a perfect fit. *Essays in Biochemistry [Online]* **2017,** *61* (5), 431–437.
55. Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A powerful approach for structure-based drug discovery. *Current computer-aided drug design [Online]* **2011,** *7* (2), 146–157.
56. Peters, J.-U. Polypharmacology - foe or friend? *Journal of medicinal chemistry [Online]* **2013,** *56* (22), 8955–8971.
57. Antolin, A. A.; Workman, P.; Mestres, J.; Al-Lazikani, B. Polypharmacology in Precision Oncology: Current Applications and Future Prospects. *Current Pharmaceutical Design [Online]* **20,** *22* (46), 6935–6945.
58. Tamkun, J. W.; Deuring, R.; Scott, M. P.; Kissinger, M.; Pattatucci, A. M.; Kaufman, T. C.; Kennison, J. A. brahma: A regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2SWI2. *Cell [Online]* **1992,** *68* (3), 561–572.
59. Galdeano, C.; Ciulli, A. Selectivity on-target of bromodomain chemical probes by structure-guided medicinal chemistry and chemical biology. *Future medicinal chemistry [Online]* **2016,** *8* (13), 1655–1680.
60. Morinière, J.; Rousseaux, S.; Steuerwald, U.; Soler-López, M.; Curtet, S.; Vitte, A.-L.; Govin, J.; Gaucher, J.; Sadoul, K.; Hart, D. J.; Krijgsveld, J.; Khochbin, S.; Müller, C. W.; Petosa, C. Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature [Online]* **2009,** *461* (7264), 664–668.
61. Robert, X.; Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research [Online]* **2014,** *42* (Web Server issue), W320-4.

62. Moustakim, M.; Clark, P. G. K.; Trulli, L.; Fuentes de Arriba, A. L.; Ehebauer, M. T.; Chaikuad, A.; Murphy, E. J.; Mendez-Johnson, J.; Daniels, D.; Hou, C.-F. D.; Lin, Y.-H.; Walker, J. R.; Hui, R.; Yang, H.; Dorrell, L.; Rogers, C. M.; Monteiro, O. P.; Fedorov, O.; Huber, K. V. M.; Knapp, S.; Heer, J.; Dixon, D. J.; Brennan, P. E. Discovery of a PCAF Bromodomain Chemical Probe. *Angewandte Chemie (International ed. in English) [Online]* **2017,** *56* (3), 827–831.

63. Humphreys, P. G.; Bamborough, P.; Chung, C.-W.; Craggs, P. D.; Gordon, L.; Grandi, P.; Hayhow, T. G.; Hussain, J.; Jones, K. L.; Lindon, M.; Michon, A.-M.; Renaux, J. F.; Suckling, C. J.; Tough, D. F.; Prinjha, R. K. Discovery of a Potent, Cell Penetrant, and Selective p300/CBP-Associated Factor (PCAF)/General Control Nonderepressible 5 (GCN5) Bromodomain Chemical Probe. *Journal of medicinal chemistry [Online]* **2017,** *60* (2), 695–709.

64. Footz, T. K.; Brinkman-Mills, P.; Banting, G. S.; Maier, S. A.; Riazi, M. A.; Bridgland, L.; Hu, S.; Birren, B.; Minoshima, S.; Shimizu, N.; Pan, H.; Nguyen, T.; Fang, F.; Fu, Y.; Ray, L.; Wu, H.; Shaull, S.; Phan, S.; Yao, Z.; Chen, F.; Huan, A.; Hu, P.; Wang, Q.; Loh, P.; Qi, S.; Roe, B. A.; McDermid, H. E. Analysis of the Cat Eye Syndrome Critical Region in Humans and the Region of Conserved Synteny in Mice: A Search for Candidate Genes at or near the Human Chromosome 22 Pericentromere. *Genome Research [Online]* **2001,** *11* (6), 1053–1070.

65. Crawford, T. D.; Audia, J. E.; Bellon, S.; Burdick, D. J.; Bommi-Reddy, A.; Côté, A.; Cummings, R. T.; Duplessis, M.; Flynn, E. M.; Hewitt, M.; Huang, H.-R.; Jayaram, H.; Jiang, Y.; Joshi, S.; Kiefer, J. R.; Murray, J.; Nasveschuk, C. G.; Neiss, A.; Pardo, E.; Romero, F. A.; Sandy, P.; Sims, R. J.; Tang, Y.; Taylor, A. M.; Tsui, V.; Wang, J.; Wang, S.; Wang, Y.; Xu, Z.; Zawadzke, L.; Zhu, X.; Albrecht, B. K.; Magnuson, S. R.; Cochran, A. G. GNE-886: A Potent and Selective Inhibitor of the Cat Eye Syndrome Chromosome Region Candidate 2 Bromodomain (CECR2). *ACS medicinal chemistry letters [Online]* **2017,** *8* (7), 737–741.

66. Jennings, L. E.; Schiedel, M.; Hewings, D. S.; Picaud, S.; Laurin, C. M. C.; Bruno, P. A.; Bluck, J. P.; Scorah, A. R.; See, L.; Reynolds, J. K.; Moroglu, M.; Mistry, I. N.; Hicks, A.; Guzanov, P.; Clayton, J.; Evans, C. N. G.; Stazi, G.; Biggin, P. C.; Mapp, A. K.; Hammond, E. M.; Humphreys, P. G.; Filippakopoulos, P.; Conway, S. J. BET bromodomain ligands: Probing the WPF shelf to improve BRD4 bromodomain affinity and metabolic stability. *Bioorganic & medicinal chemistry [Online]* **2018,** *26* (11), 2937–2957.

67. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic acids research [Online]* **2018**.

68. Mohan, M.; Lin, C.; Guest, E.; Shilatifard, A. Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. *Nature reviews. Cancer [Online]* **2010,** *10* (10), 721–728.

69. Ackloo, S.; Brown, P. J.; Müller, S. Chemical probes targeting epigenetic proteins: Applications beyond oncology. *Epigenetics [Online]* **2017,** *12* (5), 378–400.

70. Matzuk, M. M.; McKeown, M. R.; Filippakopoulos, P.; Li, Q.; Ma, L.; Agno, J. E.; Lemieux, M. E.; Picaud, S.; Yu, R. N.; Qi, J.; Knapp, S.; Bradner, J. E. Small-molecule inhibition of BRDT for male contraception. *Cell [Online]* **2012,** *150* (4), 673–684.

71. Moustakim, M.; Clark, P. G. K.; Hay, D. A.; Dixon, D. J.; Brennan, P. E. Chemical probes and inhibitors of bromodomains outside the BET family. *MedChemComm [Online]* **2016,** *7* (12), 2246–2264.

72. Hay, D. A.; Fedorov, O.; Martin, S.; Singleton, D. C.; Tallant, C.; Wells, C.; Picaud, S.; Philpott, M.; Monteiro, O. P.; Rogers, C. M.; Conway, S. J.; Rooney, T. P. C.; Tumber, A.; Yapp, C.; Filippakopoulos, P.; Bunnage, M. E.; Müller, S.; Knapp, S.; Schofield, C. J.; Brennan, P. E. Discovery and optimization of small-molecule ligands for the CBP/p300 bromodomains. *Journal of the American Chemical Society [Online]* **2014,** *136* (26), 9308–9319.

73. Hammitzsch, A.; Tallant, C.; Fedorov, O.; O'Mahony, A.; Brennan, P. E.; Hay, D. A.; Martinez, F. O.; Al-Mossawi, M. H.; Wit, J. de; Vecellio, M.; Wells, C.; Wordsworth, P.; Müller, S.; Knapp, S.; Bowness, P. CBP30, a selective CBP/p300 bromodomain inhibitor, suppresses human Th17 responses. *Proceedings of the National Academy of Sciences of the United States of America [Online]* **2015,** *112* (34), 10768–10773.

74. Picaud, S.; Fedorov, O.; Thanasopoulou, A.; Leonards, K.; Jones, K.; Meier, J.; Olzscha, H.; Monteiro, O.; Martin, S.; Philpott, M.; Tumber, A.; Filippakopoulos, P.; Yapp, C.; Wells, C.; Che, K. H.; Bannister, A.; Robson, S.; Kumar, U.; Parr, N.; Lee, K.; Lugo, D.; Jeffrey, P.; Taylor, S.; Vecellio, M. L.; Bountra, C.; Brennan, P. E.; O'Mahony, A.; Velichko, S.; Müller, S.; Hay, D.; Daniels, D. L.; Urh, M.; La Thangue, N. B.; Kouzarides, T.; Prinjha, R.; Schwaller, J.; Knapp, S. Generation of a Selective Small Molecule Inhibitor of the CBP/p300 Bromodomain for Leukemia Therapy. *Cancer research [Online]* **2015,** *75* (23), 5106–5119.

75. Romero, F. A.; Murray, J.; Lai, K. W.; Tsui, V.; Albrecht, B. K.; Le An; Beresini, M. H.; Leon Boenig, G. de; Bronner, S. M.; Chan, E. W.; Chen, K. X.; Chen, Z.; Choo, E. F.; Clagg, K.; Clark, K.; Crawford, T. D.; Cyr, P.; Almeida Nagata, D. de; Gascoigne, K. E.; Grogan, J. L.; Hatzivassiliou, G.; Huang, W.; Hunsaker, T. L.; Kaufman, S.; Koenig, S. G.; Li, R.; Li, Y.; Liang, X.; Liao, J.; Liu, W.; Ly, J.; Maher, J.; Masui, C.; Merchant, M.; Ran, Y.; Taylor, A. M.; Wai, J.; Wang, F.; Wei, X.; Yu, D.; Zhu, B.-Y.; Zhu, X.; Magnuson, S. GNE-781, A Highly Advanced Potent and Selective Bromodomain Inhibitor of Cyclic Adenosine Monophosphate Response Element Binding Protein, Binding Protein (CBP). *Journal of medicinal chemistry [Online]* **2017,** *60* (22), 9162–9183.

76. Lai, K. W.; Romero, F. A.; Tsui, V.; Beresini, M. H.; Leon Boenig, G. de; Bronner, S. M.; Chen, K.; Chen, Z.; Choo, E. F.; Crawford, T. D.; Cyr, P.; Kaufman, S.; Li, Y.; Liao, J.; Liu, W.; Ly, J.; Murray, J.; Shen, W.; Wai, J.; Wang, F.; Zhu, C.; Zhu, X.; Magnuson, S. Design and synthesis of a biaryl series as inhibitors for the bromodomains of CBP/P300. *Bioorganic & medicinal chemistry letters [Online]* **2018,** *28* (1), 15–23.

77. Martin, L. J.; Koegl, M.; Bader, G.; Cockcroft, X.-L.; Fedorov, O.; Fiegen, D.; Gerstberger, T.; Hofmann, M. H.; Hohmann, A. F.; Kessler, D.; Knapp, S.; Knesl, P.; Kornigg, S.; Müller, S.; Nar, H.; Rogers, C.; Rumpel, K.; Schaaf, O.; Steurer, S.; Tallant, C.; Vakoc, C. R.; Zeeb, M.; Zoephel, A.; Pearson, M.; Boehmelt, G.; McConnell, D. Structure-Based Design of an in Vivo Active Selective BRD9 Inhibitor. *Journal of medicinal chemistry [Online]* **2016,** *59* (10), 4462–4475.

78. Clark, P. G. K.; Vieira, L. C. C.; Tallant, C.; Fedorov, O.; Singleton, D. C.; Rogers, C. M.; Monteiro, O. P.; Bennett, J. M.; Baronio, R.; Müller, S.; Daniels, D. L.; Méndez, J.; Knapp, S.; Brennan, P. E.; Dixon, D. J. LP99: Discovery and Synthesis of the First Selective BRD7/9 Bromodomain Inhibitor. *Angewandte Chemie (International ed. in English) [Online]* **2015,** *54* (21), 6217–6221.

79. Theodoulou, N. H.; Bamborough, P.; Bannister, A. J.; Becher, I.; Bit, R. A.; Che, K. H.; Chung, C.-W.; Dittmann, A.; Drewes, G.; Drewry, D. H.; Gordon, L.; Grandi, P.; Leveridge, M.; Lindon, M.; Michon, A.-M.; Molnar, J.; Robson, S. C.; Tomkinson, N. C. O.; Kouzarides, T.; Prinjha, R. K.; Humphreys, P. G.

Discovery of I-BRD9, a Selective Cell Active Chemical Probe for Bromodomain Containing Protein 9 Inhibition. *Journal of medicinal chemistry [Online]* **2016,** *59* (4), 1425–1439.

80. Bamborough, P.; Chung, C.-W.; Demont, E. H.; Furze, R. C.; Bannister, A. J.; Che, K. H.; Diallo, H.; Douault, C.; Grandi, P.; Kouzarides, T.; Michon, A.-M.; Mitchell, D. J.; Prinjha, R. K.; Rau, C.; Robson, S.; Sheppard, R. J.; Upton, R.; Watson, R. J. A Chemical Probe for the ATAD2 Bromodomain. *Angewandte Chemie (International ed. in English) [Online]* **2016,** *55* (38), 11382–11386.

81. Fernández-Montalván, A. E.; Berger, M.; Kuropka, B.; Koo, S. J.; Badock, V.; Weiske, J.; Puetter, V.; Holton, S. J.; Stöckigt, D.; Ter Laak, A.; Centrella, P. A.; Clark, M. A.; Dumelin, C. E.; Sigel, E. A.; Soutter, H. H.; Troast, D. M.; Zhang, Y.; Cuozzo, J. W.; Keefe, A. D.; Roche, D.; Rodeschini, V.; Chaikuad, A.; Díaz-Sáez, L.; Bennett, J. M.; Fedorov, O.; Huber, K. V. M.; Hübner, J.; Weinmann, H.; Hartung, I. V.; Gorjánácz, M. Isoform-Selective ATAD2 Chemical Probe with Novel Chemical Structure and Unusual Mode of Action. *ACS chemical biology [Online]* **2017,** *12* (11), 2730–2736.

82. Meier, J. C.; Tallant, C.; Fedorov, O.; Witwicka, H.; Hwang, S.-Y.; van Stiphout, R. G.; Lambert, J.-P.; Rogers, C.; Yapp, C.; Gerstenberger, B. S.; Fedele, V.; Savitsky, P.; Heidenreich, D.; Daniels, D. L.; Owen, D. R.; Fish, P. V.; Igoe, N. M.; Bayle, E. D.; Haendler, B.; Oppermann, U. C. T.; Buffa, F.; Brennan, P. E.; Müller, S.; Gingras, A. C.; Odgren, P. R.; Birnbaum, M. J.; Knapp, S. Selective Targeting of Bromodomains of the Bromodomain-PHD Fingers Family Impairs Osteoclast Differentiation. *ACS chemical biology [Online]* **2017,** *12* (10), 2619–2630.

83. Bamborough, P.; Barnett, H. A.; Becher, I.; Bird, M. J.; Chung, C.-W.; Craggs, P. D.; Demont, E. H.; Diallo, H.; Fallon, D. J.; Gordon, L. J.; Grandi, P.; Hobbs, C. I.; Hooper-Greenhill, E.; Jones, E. J.; Law, R. P.; Le Gall, A.; Lugo, D.; Michon, A.-M.; Mitchell, D. J.; Prinjha, R. K.; Sheppard, R. J.; Watson, A. J. B.; Watson, R. J. GSK6853, a Chemical Probe for Inhibition of the BRPF1 Bromodomain. *ACS medicinal chemistry letters [Online]* **2016,** *7* (6), 552–557.

84. Bouché, L.; Christ, C. D.; Siegel, S.; Fernández-Montalván, A. E.; Holton, S. J.; Fedorov, O.; Ter Laak, A.; Sugawara, T.; Stöckigt, D.; Tallant, C.; Bennett, J.; Monteiro, O.; Díaz-Sáez, L.; Siejka, P.; Meier, J.; Pütter, V.; Weiske, J.; Müller, S.; Huber, K. V. M.; Hartung, I. V.; Haendler, B. Benzoisoquinolinediones as Potent and Selective Inhibitors of BRPF2 and TAF1/TAF1L Bromodomains. *Journal of medicinal chemistry [Online]* **2017,** *60* (9), 4002–4022.

85. Bennett, J.; Fedorov, O.; Tallant, C.; Monteiro, O.; Meier, J.; Gamble, V.; Savitsky, P.; Nunez-Alonso, G. A.; Haendler, B.; Rogers, C.; Brennan, P. E.; Müller, S.; Knapp, S. Discovery of a Chemical Tool Inhibitor Targeting the Bromodomains of TRIM24 and BRPF. *Journal of medicinal chemistry [Online]* **2016,** *59* (4), 1642–1647.

86. Chen, P.; Chaikuad, A.; Bamborough, P.; Bantscheff, M.; Bountra, C.; Chung, C.-W.; Fedorov, O.; Grandi, P.; Jung, D.; Lesniak, R.; Lindon, M.; Müller, S.; Philpott, M.; Prinjha, R.; Rogers, C.; Selenski, C.; Tallant, C.; Werner, T.; Willson, T. M.; Knapp, S.; Drewry, D. H. Discovery and Characterization of GSK2801, a Selective Chemical Probe for the Bromodomains BAZ2A and BAZ2B. *Journal of medicinal chemistry [Online]* **2016,** *59* (4), 1410–1424.

87. Drouin, L.; McGrath, S.; Vidler, L. R.; Chaikuad, A.; Monteiro, O.; Tallant, C.; Philpott, M.; Rogers, C.; Fedorov, O.; Liu, M.; Akhtar, W.; Hayes, A.; Raynaud, F.; Müller, S.; Knapp, S.; Hoelder, S. Structure enabled design of BAZ2-ICR, a chemical probe targeting the bromodomains of BAZ2A and BAZ2B. *Journal of medicinal chemistry [Online]* **2015,** *58* (5), 2553–2559.

88. Masliah-Planchon, J.; Bièche, I.; Guinebretière, J.-M.; Bourdeaut, F.; Delattre, O. SWI/SNF chromatin remodeling and human malignancies. *Annual review of pathology [Online]* **2015,** *10*, 145–171.

89. Fedorov, O.; Castex, J.; Tallant, C.; Owen, D. R.; Martin, S.; Aldeghi, M.; Monteiro, O.; Filippakopoulos, P.; Picaud, S.; Trzupek, J. D.; Gerstenberger, B. S.; Bountra, C.; Willmann, D.; Wells, C.; Philpott, M.; Rogers, C.; Biggin, P. C.; Brennan, P. E.; Bunnage, M. E.; Schüle, R.; Günther, T.; Knapp, S.; Müller, S. Selective targeting of the BRG/PB1 bromodomains impairs embryonic and trophoblast stem cell maintenance. *Science advances [Online]* **2015,** *1* (10), e1500723.

90. Allfrey, V. G.; Faulkner, R.; Mirsky, A. E. ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS*. *Proceedings of the National Academy of Sciences of the United States of America [Online]* **1964,** *51* (5), 786–794.

91. Doyon, Y.; Selleck, W.; Lane, W. S.; Tan, S.; Côté, J. Structural and Functional Conservation of the NuA4 Histone Acetyltransferase Complex from Yeast to Humans. *Molecular and Cellular Biology [Online]* **2004,** *24* (5), 1884–1896.

92. Hsu, C.-C.; Zhao, D.; Shi, J.; Peng, D.; Guan, H.; Li, Y.; Huang, Y.; Wen, H.; Li, W.; Li, H.; Shi, X. Gas41 links histone acetylation to H2A.Z deposition and maintenance of embryonic stem cell identity. *Cell Discovery [Online]***,** *4* (1), 28. https://www.nature.com/articles/s41421-018-0027-0.pdf.

93. Hetzner, K.; Garcia-Cuellar, M.-P.; Büttner, C.; Slany, R. K. The interaction of ENL with PAF1 mitigates polycomb silencing and facilitates murine leukemogenesis. *Blood [Online]* **2018,** *131* (6), 662–673.

94. He, N.; Chan, C. K.; Sobhian, B.; Chou, S.; Xue, Y.; Liu, M.; Alber, T.; Benkirane, M.; Zhou, Q. Human Polymerase-Associated Factor complex (PAFc) connects the Super Elongation Complex (SEC) to RNA polymerase II on chromatin. *Proceedings of the National Academy of Sciences of the United States of America [Online]* **2011,** *108* (36), E636-45.

95. Nie, Z.; Yan, Z.; Chen, E. H.; Sechi, S.; Ling, C.; Zhou, S.; Xue, Y.; Yang, D.; Murray, D.; Kanakubo, E.; Cleary, M. L.; Wang, W. Novel SWI/SNF Chromatin-Remodeling Complexes Contain a Mixed-Lineage Leukemia Chromosomal Translocation Partner. *Molecular and Cellular Biology [Online]* **2003,** *23* (8), 2942–2952.

96. Perlman, E. J.; Gadd, S.; Arold, S. T.; Radhakrishnan, A.; Gerhard, D. S.; Jennings, L.; Huff, V.; Guidry Auvil, J. M.; Davidsen, T. M.; Dome, J. S.; Meerzaman, D.; Hsu, C. H.; Nguyen, C.; Anderson, J.; Ma, Y.; Mungall, A. J.; Moore, R. A.; Marra, M. A.; Mullighan, C. G.; Ma, J.; Wheeler, D. A.; Hampton, O. A.; Gastier-Foster, J. M.; Ross, N.; Smith, M. A. MLLT1 YEATS domain mutations in clinically distinctive Favourable Histology Wilms tumours. *Nature Communications [Online]* **2015,** *6*, 10013.

97. Mi, W.; Guan, H.; Lyu, J.; Zhao, D.; Xi, Y.; Jiang, S.; Andrews, F. H.; Wang, X.; Gagea, M.; Wen, H.; Tora, L.; Dent, S. Y. R.; Kutateladze, T. G.; Li, W.; Li, H.; Shi, X. YEATS2 links histone acetylation to tumorigenesis of non-small cell lung cancer. *Nature Communications [Online]***,** *8* (1), 1088. https://www.nature.com/articles/s41467-017-01173-4.pdf.

98. Guelman, S.; Kozuka, K.; Mao, Y.; Pham, V.; Solloway, M. J.; Wang, J.; Wu, J.; Lill, J. R.; Zha, J. The double-histone-acetyltransferase complex ATAC is essential for mammalian development. *Molecular and Cellular Biology [Online]* **2009,** *29* (5), 1176–1188.

99. Wong, M. M.; Cox, L. K.; Chrivia, J. C. The Chromatin Remodeling Protein, SRCAP, Is Critical for Deposition of the Histone Variant H2A.Z at Promoters. *J. Biol. Chem. [Online]* **2007,** *282* (36), 26132–26139. http://www.jbc.org/content/282/36/26132.full.

100. Fischer, U.; Meltzer, P.; Meese, E. Twelve amplified and expressed genes localized in a single domain in glioma. *Human genetics [Online]* **1996,** *98* (5), 625–628.

101. Tao, K.; Yang, J.; Hu, Y.; Deng, A. Knockdown of YEATS4 inhibits colorectal cancer cell proliferation and induces apoptosis. *American journal of translational research [Online]* **2015,** *7* (3), 616–623.

102. Jixiang, C.; Shengchun, D.; Jianguo, Q.; Zhengfa, M.; Xin, F.; Xuqing, W.; Jianxin, Z.; Lei, C. YEATS4 promotes the tumorigenesis of pancreatic cancer by activating beta-catenin/TCF signaling. *Oncotarget [Online]* **2017,** *8* (15), 25200–25210.

103. Heisel, S.; Habel, N. C.; Schuetz, N.; Ruggieri, A.; Meese, E. The YEATS family member GAS41 interacts with the general transcription factor TFIIF. *BMC molecular biology [Online]* **2010,** *11*, 53.

104. Wilkinson, A. W.; Gozani, O. Cancer epigenetics: Reading the future of leukaemia. *Nature [Online]* **2017,** *543* (7644), 186–188.

105. Park, J. R.; Eggert, A.; Caron, H. Neuroblastoma: biology, prognosis, and treatment. *Pediatric clinics of North America [Online]* **2008,** *55* (1), 97-120, x.

106. Berry, T.; Luther, W.; Bhatnagar, N.; Jamin, Y.; Poon, E.; Sanda, T.; Pei, D.; Sharma, B.; Vetharoy, W. R.; Hallsworth, A.; Ahmad, Z.; Barker, K.; Moreau, L.; Webber, H.; Wang, W.; Liu, Q.; Perez-Atayde, A.; Rodig, S.; Cheung, N.-K.; Raynaud, F.; Hallberg, B.; Robinson, S. P.; Gray, N. S.; Pearson, A. D.J.; Eccles, S. A.; Chesler, L.; George, R. E. The ALKF1174L mutation potentiates the oncogenic activity of MYCN in neuroblastoma. *Cancer cell [Online]* **2012,** *22* (1), 117–130.

107. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography [Online]* **2011,** *67* (Pt 4), 235–242.

108. Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Cryst D [Online]* **2004,** *60* (Pt 12 Pt 1), 2126–2132.

109. Linding, R.; Russell, R. B.; Neduva, V.; Gibson, T. J. GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic acids research [Online]* **2003,** *31* (13), 3701–3708.

110. Drozdetskiy, A.; Cole, C.; Procter, J.; Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic acids research [Online]* **2015,** *43* (W1), W389-94.

111. Aslanidis, C.; Jong, P. J. de. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic acids research [Online]* **1990,** *18* (20), 6069–6074.

112. Saiki, R. K.; Gelfand, D. H.; Stoffel, S.; Scharf, S. J.; Higuchi, R.; Horn, G. T.; Mullis, K. B.; Erlich, H. A. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science (New York, N.Y.) [Online]* **1988,** *239* (4839), 487–491.

113. Gay, P.; Le Coq, D.; Steinmetz, M.; Berkelman, T.; Kado, C. I. Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *Journal of Bacteriology [Online]* **1985,** *164* (2), 918–921.

114. Recorbet, G.; Robert, C.; Givaudan, A.; Kudla, B.; Normand, P.; Faurie, G. Conditional suicide system of Escherichia coli released into soil that uses the Bacillus subtilis sacB gene. *Applied and Environmental Microbiology [Online]* **1993,** *59* (5), 1361–1366.

115. Mandel, M.; Higa, A. Calcium-dependent bacteriophage DNA infection. *Journal of molecular biology [Online]* **1970,** *53* (1), 159–162.

116. Hanahan, D. Studies on transformation of Escherichia coli with plasmids. *Journal of molecular biology [Online]* **1983,** *166* (4), 557–580.

117. Tabor, S.; Richardson, C. C. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proceedings of the National Academy of Sciences of the United States of America [Online]* **1985,** *82* (4), 1074–1078.

118. Studier, F. W.; Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology [Online]* **1986,** *189* (1), 113–130.

119. Makrides, S. C. Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiological Reviews [Online]* **1996,** *60* (3), 512–538.

120. Lindqvist, B.; Storgards, T. Molecular-sieving Properties of Starch. *Nature [Online]* **1955,** *175*, 511 EP -. https://doi.org/10.1038/175511a0.

121. Porath, J.; Flodin, P. Gel filtration: a method for desalting and group separation. *Nature [Online]* **1959,** *183* (4676), 1657–1659.

122. Gill, S. C.; Hippel, P. H. von. Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry [Online]* **1989,** *182* (2), 319–326.

123. Beer, A. Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. *Annalen der Physik und Chemie [Online]* **1852,** No. 86, 78–88.

124. Anthis, N. J.; Clore, G. M. Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Science : A Publication of the Protein Society [Online]* **2013,** *22* (6), 851–858.

125. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature [Online]* **1970,** *227* (5259), 680–685.

126. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science (New York, N.Y.) [Online]* **1989,** *246* (4926), 64–71.

127. Iribarne, J. V. On the evaporation of small ions from charged droplets. *J. Chem. Phys. [Online]* **1976,** *64* (6), 2287.

128. Baldwin, M. A. Mass spectrometers for the analysis of biomolecules. *Methods in enzymology [Online]* **2005,** *402*, 3–48.

129. El-Aneed, A.; Cohen, A.; Banoub, J. Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Applied Spectroscopy Reviews [Online]* **2009,** *44* (3), 210–230.

130. Niesen, F. H.; Berglund, H.; Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature protocols [Online]* **2007,** *2* (9), 2212–2221.

131. Freyer, M. W.; Lewis, E. A. Isothermal Titration Calorimetry: Experimental Design, Data Analysis, and Probing Macromolecule/Ligand Binding and Kinetic Interactions. In *Biophysical tools for biologists*: *Volume 1: In vitro techniques*; Detrich, H. W., Correia, J. J., Eds.; Methods in Cell Biology v. 84; Elsevier: Amsterdam, s.l., 2008; pp 79–113.

132. 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 chemical biology [Online]* **2015,** *10* (8), 1797–1804.

133. 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 chemical biology [Online]* **2018,** *25* (2), 206-214.e11.

134. Hall, M. P.; Unch, J.; Binkowski, B. F.; Valley, M. P.; Butler, B. L.; Wood, M. G.; Otto, P.; Zimmerman, K.; Vidugiris, G.; Machleidt, T.; Robers, M. B.; Benink, H. A.; Eggers, C. T.; Slater, M. R.; Meisenheimer, P. L.; Klaubert, D. H.; Fan, F.; Encell, L. P.; Wood, K. V. Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS chemical biology [Online]* **2012,** *7* (11), 1848–1857.

135. Rhodes, G. *Crystallography made crystal clear*. *A guide for users of macromolecular models*, 3rd ed.; Complementary science series; Elsevier/Academic Press: Amsterdam, Boston, 2006.

136. McPherson, A.; Gavira, J. A. Introduction to protein crystallization. *Acta crystallographica. Section F, Structural biology communications [Online]* **2014,** *70* (Pt 1), 2–20.

137. Krojer, T.; Pike, A. C. W.; Delft, F. von. Squeezing the most from every crystal: the fine details of data collection. *Acta crystallographica. Section D, Biological crystallography [Online]* **2013,** *69* (Pt 7), 1303–1313.

138. Powell, H. R.; Battye, T. G. G.; Kontogiannis, L.; Johnson, O.; Leslie, A. G. W. Integrating macromolecular X-ray diffraction data with the graphical user interface iMosflm. *Nature protocols [Online]* **2017,** *12* (7), 1310–1325.

139. Krug, M.; Weiss, M. S.; Heinemann, U.; Mueller, U. XDSAPP : a graphical user interface for the convenient processing of diffraction data using XDS. *J Appl Crystallogr [Online]* **2012,** *45* (3), 568–572.

140. Sparta, K. M.; Krug, M.; Heinemann, U.; Mueller, U.; Weiss, M. S. XDSAPP2.0. *J Appl Crystallogr [Online]* **2016,** *49* (3), 1085–1092.

141. Powell, H. R. X-ray data processing. *Bioscience reports* **2017,** *37* (5). DOI: 10.1042/BSR20170227.

142. Evans, P. R.; Murshudov, G. N. How good are my data and what is the resolution? *Acta crystallographica. Section D, Biological crystallography [Online]* **2013,** *69* (Pt 7), 1204–1214.

143. Karplus, P. A.; Diederichs, K. Linking crystallographic model and data quality. *Science (New York, N.Y.) [Online]* **2012,** *336* (6084), 1030–1033.

144. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *Journal of applied crystallography [Online]* **2007,** *40* (Pt 4), 658–674.

145. McCoy, A. J. Liking likelihood. *Acta Cryst D [Online]* **2004,** *60* (12), 2169–2183. //journals.iucr.org/d/issues/2004/12/01/ba5064/ba5064.pdf.

146. Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst D [Online]* **2011,** *67* (Pt 4), 355–367.

147. Terwilliger, T. C. Reciprocal-space solvent flattening. *Acta Cryst D [Online]* **1999,** *55* (11), 1863–1871.

148. Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *Journal of molecular biology [Online]* **1963,** *7* (1), 95–99.

149. Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography [Online]* **2010,** *66* (Pt 1), 12–21.

150. Painter, J.; Merritt, E. A. TLSMD web server for the generation of multi-group TLS models. *J Appl Crystallogr [Online]* **2006,** *39* (1), 109–111.

151. Miller, D. M.; Thomas, S. D.; Islam, A.; Muench, D.; Sedoris, K. c-Myc and Cancer Metabolism. *Clinical cancer research : an official journal of the American Association for Cancer Research [Online]* **2012,** *18* (20), 5546–5553.

152. Bresler, S. C.; Wood, A. C.; Haglund, E. A.; Courtright, J.; Belcastro, L. T.; Plegaria, J. S.; Cole, K.; Toporovskaya, Y.; Zhao, H.; Carpenter, E. L.; Christensen, J. G.; Maris, J. M.; Lemmon, M. A.; Mossé, Y. P. Differential inhibitor sensitivity of anaplastic lymphoma kinase variants found in neuroblastoma. *Science translational medicine [Online]* **2011,** *3* (108), 108ra114.

153. Guan, J.; Tucker, E. R.; Wan, H.; Chand, D.; Danielson, L. S.; Ruuth, K.; El Wakil, A.; Witek, B.; Jamin, Y.; Umapathy, G.; Robinson, S. P.; Johnson, T. W.; Smeal, T.; Martinsson, T.; Chesler, L.; Palmer, R. H.; Hallberg, B. The ALK inhibitor PF-06463922 is effective as a single agent in neuroblastoma driven by expression of ALK and MYCN. *Disease Models & Mechanisms [Online]* **2016,** *9* (9), 941–952.

154. Chaikuad, A.; Petros, A. M.; Fedorov, O.; Xu, J.; Knapp, S. Structure-based approaches towards identification of fragments for the low-druggability ATAD2 bromodomain. *Med. Chem. Commun. [Online]* **2014,** *5* (12), 1843–1848. https://pubs.rsc.org/en/content/articlepdf/2014/md/c4md00237g.

155. Chaikuad, A.; Lang, S.; Brennan, P. E.; Temperini, C.; Fedorov, O.; Hollander, J.; Nachane, R.; Abell, C.; Müller, S.; Siegal, G.; Knapp, S. Structure-Based Identification of Inhibitory Fragments Targeting the p300/CBP-Associated Factor Bromodomain. *Journal of medicinal chemistry [Online]* **2016,** *59* (4), 1648–1653.

156. Shang, J.; Dai, X.; Li, Y.; Pistolozzi, M.; Wang, L. HybridSim-VS: a web server for large-scale ligand-based virtual screening using hybrid similarity recognition techniques. *Bioinformatics (Oxford, England) [Online]* **2017,** *33* (21), 3480–3481.

157. Pèpe, G.; Guiliani, G.; Loustalet, S.; Halfon, P. Hydration free energy a fragmental model and drug design. *European journal of medicinal chemistry [Online]* **2002,** *37* (11), 865–872.

158. Nittinger, E.; Gibbons, P.; Eigenbrot, C.; Davies, D. R.; Maurer, B.; Yu, C. L.; Kiefer, J. R.; Kuglstatter, A.; Murray, J.; Ortwine, D. F.; Tang, Y.; Tsui, V. Water molecules in protein-ligand interfaces. Evaluation of software tools and SAR comparison. *Journal of computer-aided molecular design [Online]* **2019**.

159. Wasylyk, C.; Schlumberger, S. E.; Criqui-Filipe, P.; Wasylyk, B. Sp100 interacts with ETS-1 and stimulates its transcriptional activity. *Molecular and Cellular Biology [Online]* **2002,** *22* (8), 2687–2702.

160. Bernardi, R.; Pandolfi, P. P. Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nature reviews. Molecular cell biology [Online]* **2007,** *8* (12), 1006–1016.

161. Meroni, G.; Diez-Roux, G. TRIM/RBCC, a novel class of 'single protein RING finger' E3 ubiquitin ligases. *BioEssays : news and reviews in molecular, cellular and developmental biology [Online]* **2005,** *27* (11), 1147–1157.

162. Liu, J.; Li, F.; Bao, H.; Jiang, Y.; Zhang, S.; Ma, R.; Gao, J.; Wu, J.; Ruan, K. The polar warhead of a TRIM24 bromodomain inhibitor rearranges a water-mediated interaction network. *The FEBS journal [Online]* **2017,** *284* (7), 1082–1095.

163. Slaughter, M. J.; Shanle, E. K.; McFadden, A. W.; Hollis, E. S.; Suttle, L. E.; Strahl, B. D.; Davis, I. J. PBRM1 bromodomains variably influence nucleosome interactions and cellular function. *J. Biol. Chem. [Online]* **2018,** *293* (35), 13592–13603.

164. Porter, E. G.; Dykhuizen, E. C. Individual Bromodomains of Polybromo-1 Contribute to Chromatin Association and Tumor Suppression in Clear Cell Renal Carcinoma. *J. Biol. Chem. [Online]* **2017,** *292* (7), 2601–2610.

165. Liao, L.; Alicea-Velázquez, N. L.; Langbein, L.; Niu, X.; Cai, W.; Cho, E.-A.; Zhang, M.; Greer, C. B.; Yan, Q.; Cosgrove, M. S.; Yang, H. High affinity binding of H3K14ac through collaboration of bromodomains 2, 4 and 5 is critical for the molecular and tumor suppressor functions of PBRM1. *Molecular oncology [Online]* **2019,** *13* (4), 811–828.

166. Cherrier, T.; Le Douce, V.; Eilebrecht, S.; Riclet, R.; Marban, C.; Dequiedt, F.; Goumon, Y.; Paillart, J.-C.; Mericskay, M.; Parlakian, A.; Bausero, P.; Abbas, W.; Herbein, G.; Kurdistani, S. K.; Grana, X.; van Driessche, B.; Schwartz, C.; Candolfi, E.; Benecke, A. G.; van Lint, C.; Rohr, O. CTIP2 is a negative regulator of P-TEFb. *Proceedings of the National Academy of Sciences of the United States of America [Online]* **2013,** *110* (31), 12655–12660.

167. Roberts, T. C.; Etxaniz, U.; Dall'Agnese, A.; Wu, S.-Y.; Chiang, C.-M.; Brennan, P. E.; Wood, M. J. A.; Puri, P. L. BRD3 and BRD4 BET Bromodomain Proteins Differentially Regulate Skeletal Myogenesis. *Scientific reports [Online]* **2017,** *7* (1), 6153.

168. Zhang, L.; Hao, C.; Dong, G.; Tong, Z. Analysis of Clinical Features and Outcome of 356 Triple-Negative Breast Cancer Patients in China. *Breast Care [Online]* **2012,** *7* (1), 13–17.

169. da Motta, L. L.; Ledaki, I.; Purshouse, K.; Haider, S.; Bastiani, M. A. de; Baban, D.; Morotti, M.; Steers, G.; Wigfield, S.; Bridges, E.; Li, J.-L.; Knapp, S.; Ebner, D.; Klamt, F.; Harris, A. L.; McIntyre, A.

The BET inhibitor JQ1 selectively impairs tumour response to hypoxia and downregulates CA9 and angiogenesis in triple negative breast cancer. *Oncogene [Online]* **2017,** *36* (1), 122–132.

170. Kim, S.; Kim, D. H.; Jung, W.-H.; Koo, J. S. Metabolic phenotypes in triple-negative breast cancer. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine [Online]* **2013,** *34* (3), 1699–1712.

171. Jeon, H. M.; Kim, D. H.; Jung, W.-H.; Koo, J. S. Expression of cell metabolism-related genes in different molecular subtypes of triple-negative breast cancer. *Tumori [Online]* **2013,** *99* (4), 555–564.

172. Wellen, K. E.; Hatzivassiliou, G.; Sachdeva, U. M.; Bui, T. V.; Cross, J. R.; Thompson, C. B. ATP-citrate lyase links cellular metabolism to histone acetylation. *Science (New York, N.Y.) [Online]* **2009,** *324* (5930), 1076–1080.

173. Siebeneicher, H.; Cleve, A.; Rehwinkel, H.; Neuhaus, R.; Heisler, I.; Müller, T.; Bauser, M.; Buchmann, B. Identification and Optimization of the First Highly Selective GLUT1 Inhibitor BAY-876. *ChemMedChem [Online]* **2016,** *11* (20), 2261–2271.

174. Lalonde, M.-E.; Avvakumov, N.; Glass, K. C.; Joncas, F.-H.; Saksouk, N.; Holliday, M.; Paquet, E.; Yan, K.; Tong, Q.; Klein, B. J.; Tan, S.; Yang, X.-J.; Kutateladze, T. G.; Côté, J. Exchange of associated factors directs a switch in HBO1 acetyltransferase histone tail specificity. *Genes & development [Online]* **2013,** *27* (18), 2009–2024.

175. Dutra, L. A.; Heidenreich, D.; Silva, G. D. B. d.; Man Chin, C.; Knapp, S.; Santos, J. L. D. Dietary Compound Resveratrol Is a Pan-BET Bromodomain Inhibitor. *Nutrients* **2017,** *9* (11). DOI: 10.3390/nu9111172.

176. Singh, U. P.; Singh, N. P.; Singh, B.; Hofseth, L. J.; Price, R. L.; Nagarkatti, M.; Nagarkatti, P. S. Resveratrol (trans-3,5,4'-trihydroxystilbene) induces silent mating type information regulation-1 and down-regulates nuclear transcription factor-kappaB activation to abrogate dextran sulfate sodium-induced colitis. *The Journal of pharmacology and experimental therapeutics [Online]* **2010,** *332* (3), 829–839.

177. Huang, B.; Yang, X.-D.; Zhou, M.-M.; Ozato, K.; Chen, L.-F. Brd4 Coactivates Transcriptional Activation of NF-κB via Specific Binding to Acetylated RelA ▽. *Molecular and Cellular Biology [Online]* **2008,** *29* (5), 1375–1387.

178. Dittmann, A.; Werner, T.; Chung, C.-W.; Savitski, M. M.; Fälth Savitski, M.; Grandi, P.; Hopf, C.; Lindon, M.; Neubauer, G.; Prinjha, R. K.; Bantscheff, M.; Drewes, G. The commonly used PI3-kinase probe LY294002 is an inhibitor of BET bromodomains. *ACS chemical biology [Online]* **2014,** *9* (2), 495–502.

179. Andrews, F. H.; Singh, A. R.; Joshi, S.; Smith, C. A.; Morales, G. A.; Garlich, J. R.; Durden, D. L.; Kutateladze, T. G. Dual-activity PI3K-BRD4 inhibitor for the orthogonal inhibition of MYC to block tumor growth and metastasis. *Proceedings of the National Academy of Sciences of the United States of America [Online]* **2017,** *114* (7), E1072-E1080.

180. Chen, L.; Yap, J. L.; Yoshioka, M.; Lanning, M. E.; Fountain, R. N.; Raje, M.; Scheenstra, J. A.; Strovel, J. W.; Fletcher, S. BRD4 Structure-Activity Relationships of Dual PLK1 Kinase/BRD4 Bromodomain Inhibitor BI-2536. *ACS medicinal chemistry letters [Online]* **2015,** *6* (7), 764–769.

181. Hu, J.; Wang, Y.; Li, Y.; Xu, L.; Cao, D.; Song, S.; Damaneh, M. S.; Wang, X.; Meng, T.; Chen, Y.-L.; Shen, J.; Miao, Z.; Xiong, B. Discovery of a series of dihydroquinoxalin-2(1H)-ones as selective BET inhibitors from a dual PLK1-BRD4 inhibitor. *European journal of medicinal chemistry [Online]* **2017,** *137*, 176–195.

182. Liu, S.; Yosief, H. O.; Dai, L.; Huang, H.; Dhawan, G.; Zhang, X.; Muthengi, A. M.; Roberts, J.; Buckley, D. L.; Perry, J. A.; Wu, L.; Bradner, J. E.; Qi, J.; Zhang, W. Structure-Guided Design and Development of Potent and Selective Dual Bromodomain 4 (BRD4)/Polo-like Kinase 1 (PLK1) Inhibitors. *Journal of medicinal chemistry [Online]* **2018,** *61* (17), 7785–7795.

183. Felgenhauer, J.; Tomino, L.; Selich-Anderson, J.; Bopp, E.; Shah, N. Dual BRD4 and AURKA Inhibition Is Synergistic against MYCN-Amplified and Nonamplified Neuroblastoma. *Neoplasia (New York, N.Y.) [Online]* **2018,** *20* (10), 965–974.

184. Zhu, X.; Holmsen, E.; Park, S.; Willingham, M. C.; Qi, J.; Cheng, S.-Y. Synergistic effects of BET and MEK inhibitors promote regression of anaplastic thyroid tumors. *Oncotarget [Online]* **2018,** *9* (83), 35408–35421.

185. Kleppe, M.; Koche, R.; Zou, L.; van Galen, P.; Hill, C. E.; Dong, L.; Groote, S. de; Papalexi, E.; Hanasoge Somasundara, A. V.; Cordner, K.; Keller, M.; Farnoud, N.; Medina, J.; McGovern, E.; Reyes, J.; Roberts, J.; Witkin, M.; Rapaport, F.; Teruya-Feldstein, J.; Qi, J.; Rampal, R.; Bernstein, B. E.; Bradner, J. E.; Levine, R. L. Dual Targeting of Oncogenic Activation and Inflammatory Signaling Increases Therapeutic Efficacy in Myeloproliferative Neoplasms. *Cancer cell [Online]* **2018,** *33* (1), 29-43.e7.

186. Manzotti, G.; Ciarrocchi, A.; Sancisi, V. Inhibition of BET Proteins and Histone Deacetylase (HDACs): Crossing Roads in Cancer Therapy. *Cancers* **2019,** *11* (3). DOI: 10.3390/cancers11030304.

187. Doetzlhofer, A.; Rotheneder, H.; Lagger, G.; Koranda, M.; Kurtev, V.; Brosch, G.; Wintersberger, E.; Seiser, C. Histone deacetylase 1 can repress transcription by binding to Sp1. *Molecular and Cellular Biology [Online]* **1999,** *19* (8), 5504–5511.

188. Wang, Z.; Zang, C.; Cui, K.; Schones, D. E.; Barski, A.; Peng, W.; Zhao, K. Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell [Online]* **2009,** *138* (5), 1019–1031.

189. Shahbazi, J.; Liu, P. Y.; Atmadibrata, B.; Bradner, J. E.; Marshall, G. M.; Lock, R. B.; Liu, T. The Bromodomain Inhibitor JQ1 and the Histone Deacetylase Inhibitor Panobinostat Synergistically Reduce N-Myc Expression and Induce Anticancer Effects. *Clinical cancer research : an official journal of the American Association for Cancer Research [Online]* **2016,** *22* (10), 2534–2544.

190. Fiskus, W.; Sharma, S.; Qi, J.; Valenta, J. A.; Schaub, L. J.; Shah, B.; Peth, K.; Portier, B. P.; Rodriguez, M.; Devaraj, S. G. T.; Zhan, M.; Sheng, J.; Iyer, S. P.; Bradner, J. E.; Bhalla, K. N. Highly active combination of BRD4 antagonist and histone deacetylase inhibitor against human acute myelogenous leukemia cells. *Molecular cancer therapeutics [Online]* **2014,** *13* (5), 1142–1154.

191. Heinemann, A.; Cullinane, C.; Paoli-Iseppi, R. de; Wilmott, J. S.; Gunatilake, D.; Madore, J.; Strbenac, D.; Yang, J. Y.; Gowrishankar, K.; Tiffen, J. C.; Prinjha, R. K.; Smithers, N.; McArthur, G. A.; Hersey, P.; Gallagher, S. J. Combining BET and HDAC inhibitors synergistically induces apoptosis of melanoma and suppresses AKT and YAP signaling. *Oncotarget [Online]* **2015,** *6* (25), 21507–21521.

192. Zhang, Y.; Ishida, C. T.; Ishida, W.; Lo, S.-F. L.; Zhao, J.; Shu, C.; Bianchetti, E.; Kleiner, G.; Sanchez-Quintero, M. J.; Quinzii, C. M.; Westhoff, M.-A.; Karpel-Massler, G.; Canoll, P.; Siegelin, M. D. Combined HDAC and Bromodomain Protein Inhibition Reprograms Tumor Cell Metabolism and Elicits Synthetic Lethality in Glioblastoma. *Clinical cancer research : an official journal of the American Association for Cancer Research [Online]* **2018,** *24* (16), 3941–3954.

193. Zhang, Z.; Hou, S.; Chen, H.; Ran, T.; Jiang, F.; Bian, Y.; Zhang, D.; Zhi, Y.; Wang, L.; Zhang, L.; Li, H.; Zhang, Y.; Tang, W.; Lu, T.; Chen, Y. Targeting epigenetic reader and eraser: Rational design, synthesis and in vitro evaluation of dimethylisoxazoles derivatives as BRD4/HDAC dual inhibitors. *Bioorganic & medicinal chemistry letters [Online]* **2016,** *26* (12), 2931–2935.

194. Shao, M.; He, L.; Zheng, L.; Huang, L.; Zhou, Y.; Wang, T.; Chen, Y.; Shen, M.; Wang, F.; Yang, Z.; Chen, L. Structure-based design, synthesis and in vitro antiproliferative effects studies of novel dual BRD4/HDAC inhibitors. *Bioorganic & medicinal chemistry letters [Online]* **2017,** *27* (17), 4051–4055.

195. Amemiya, S.; Yamaguchi, T.; Hashimoto, Y.; Noguchi-Yachide, T. Synthesis and evaluation of novel dual BRD4/HDAC inhibitors. *Bioorganic & medicinal chemistry [Online]* **2017,** *25* (14), 3677–3684.

196. Pattabiraman, D. R.; McGirr, C.; Shakhbazov, K.; Barbier, V.; Krishnan, K.; Mukhopadhyay, P.; Hawthorne, P.; Trezise, A.; Ding, J.; Grimmond, S. M.; Papathanasiou, P.; Alexander, W. S.; Perkins, A. C.; Levesque, J.-P.; Winkler, I. G.; Gonda, T. J. Interaction of c-Myb with p300 is required for the induction of acute myeloid leukemia (AML) by human AML oncogenes. *Blood [Online]* **2014,** *123* (17), 2682–2690.

197. Yang, X.; Wang, J. Precision therapy for acute myeloid leukemia. *Journal of Hematology & Oncology* **2018,** *11. DOI:* 10.1186/s13045-017-0543-7.

198. Li, X.; Li, X.-M.; Jiang, Y.; Liu, Z.; Cui, Y.; Fung, K. Y.; van der Beelen, S. H. E.; Tian, G.; Wan, L.; Shi, X.; Allis, C. D.; Li, H.; Li, Y.; Li, X. D. Structure-guided development of YEATS domain inhibitors by targeting π-π-π stacking. *Nature chemical biology [Online]* **2018,** *14* (12), 1140–1149.

199. Tallant, C.; Owen, D. R.; Gerstenberger, B. S.; Fedorov, O.; Savitsky, P.; Nunez-Alonso, G.; Fonseca, M.; Krojer, T.; Filippakopoulos, P.; Delft, F. von; Arrowsmith, C. H.; Edwards, A. M.; Bountra, C.; Muller, S.; Knapp, S. *Crystal structure of the bromodomain of human BRM (SMARCA2) in complex with PFI-3 chemical probe*, **2015**.