Simulations of Conformational Changes and Enzyme-Substrate Interactions in Protein Drug Targets

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich
Physik
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Laura Tesmer

Frankfurt am Main 2022
(D 30)

Diese Arbeit wurde vom Fachbereich Physik
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Harald Appelshäuser

1. Gutachter: Prof. Dr. Gerhard Hummer

2. Gutachter: Prof. Dr. Jens Bredenbeck

Datum der Disputation:

Die vorliegende Dissertation wurde im Zeitraum von Mai 2018 bis Januar 2022 unter der Anleitung von Herrn Prof. Dr. Gerhard Hummer in der Abteilung Theoretische Biophysik am Max-Planck-Institut für Biophysik in Frankfurt am Main angefertigt.

Contents

List of Abbreviations

ACE	Angiotensin-converting enzyme
ACPYPE	Antechamber python parser interface
AMBER	Assisted model building with energy refinement
ANP	Atrial natriuretic peptide
ATP	Adenosine triphosphate
BL2	Blocking loop 2
CK1	Casein kinase 1
COVID-19	Coronavirus disease 19
CV	Collective variable
DNA	Deoxyribonucleic acid
FVIIa	Factor VIIa
GAFF	General amber force field
Gromacs	Groningen machine for chemical simulations
HREX	Hamiltonian replica exchange
ISG-15	Interferon-stimulated gene-15
K_D	Dissociation constant
LINCS	Linear constraint solver
MD	Molecular dynamics
MERS-CoV	Middle east respiratory syndrome coronavirus
NMR	Nuclear magnetic resonance
NPC2	NPC intracellular cholesterol transporter 2
PDB	Protein data bank
PLpro	Papain-like protease
PME	Particle mesh Ewald
PMF	Potential of mean force
PPC	Periodic boundary conditions
QM	Quantum mechanics

RAAS	Renin-angiotensin-aldosterone system
REMD	Replica exchange molecular dynamics
RESP	Restraint electrostatic potential
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
SARS-CoV	Severe acute respiratory syndrome coronavirus
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SCoV-PLpro	Papain-like protease of severe acute respiratory syndrome coronavirus
SCoV2-PLpro	Papain-like protease of severe acute respiratory syndrome coronavirus 2
SWISH	Sampling water interfaces through scaled hamiltonians
TF	Tissue factor
TIP3P	Transferable intermolecular potential 3 point
Ub	Ubiquitin
vdW	van der Waals
VMD	Visual molecular dynamics
WHAM	Weighted histogram analysis method

List of Tables

List of Figures

Zusammenfassung

Die Suche nach neuen Arzneimitteln ist eine schwierige, zeitaufwändige und kostenintensive Aufgabe mit einer geringen Erfolgsquote von weit unter 10%. Die hohe Misserfolgsquote in der Arzneimittelforschung regt dazu an, mehr computergestützte Methoden in dem gesamten Arzneimittelforschungsprozess, von der *Target*-Identifizierung bis hin zu klinischen Versuchen, zu verwenden. Die *Target*-Identifizierung ist der erste Schritt in diesem Prozess. Dieser beschreibt die Suche nach einem biologischen *Target*, beispielsweise einem Protein, das funktionale Relevanz bei einem Krankheitsbild hat, auch genannt Zielprotein. In dieser Forschungsphase ist wichtig, dass der molekulare Mechanismus des Zielproteins in einem Krankheitsbild verstanden werden muss. Zudem muss eine potenzielle Bindungsstelle des Zielproteins, an der therapeutische Moleküle binden und somit die Aktivität des Zielproteins modulieren können, charakterisiert werden. Computergestützte Methoden können dazu beitragen, die molekulare Zielproteinermittlung und -bewertung zu verbessern.

In dieser Arbeit verwende ich computergestützte, physikbasierte Ansätze, um die Bindungsstellen von Zielproteinen zu charakterisieren und die Wechselwirkungen zwischen Enzymen und Substraten zu entschlüsseln, die bei Krankheitsmechanismen eine Rolle spielen. Mit Hilfe von Molekulardynamiksimulationen (MD-Simulationen) wurde die Dynamik von Molekülen in Lösung mit hoher zeitlicher und räumlicher Auflösung untersucht. Ausgehend von einem Satz von Koordinaten und Geschwindigkeiten erzeugt die Methode zeitaufgelöste Trajektorien der Teilchen des untersuchten Systems durch numerische Integration der Newtonschen Bewegungsgleichungen. Bei MD-Simulationen werden alle Atome eines ausgewählten Systems, einschließlich des Lösungsmittels, explizit dargestellt. Atomistische Simulationen eignen sich besonders gut für die Untersuchung detaillierter Wechselwirkungen, die von intermolekularen Wechselwirkungen abhängen, wie Hydratationseffekte, Wasserstoffbrückenbindungen, hydrophobe Wechselwirkungen oder subtile chemische Unterschiede. Die Analyse der Trajektorien ermöglicht die Vorhersage von Systemeigenschaften, sofern die Kraftfelder, die die Wechselwirkungen zwischen den Teilchen im System beschreiben, eine hohe Genauigkeit aufweisen. Die gebundenen und nicht gebundenen Wechselwirkungen werden anhand experimenteller und quantenchemischer Daten parametrisiert. Der Zweck von MD-Simulationen kann darin bestehen, Einblicke in das Verhalten komplexer biologischer Systeme auf molekularer Ebene zu gewinnen, die in Experimenten oft nicht mit der gleichen Auflösung beobachtet werden können. Mit der jüngsten Entwicklung von Computerhardware und Simulationssoftware können molekulare Systeme mit zunehmender Größe und Simulationsdauer untersucht werden.

Im ersten Teil der Arbeit habe ich das Konformationsensemble verschiedener Zielproteine für Arzneimittel untersucht. Proteine sind dynamische Biomakromoleküle, die verschiedene und nahezu isoenergetische Konformationszustände aufweisen können. Eine Ligandenbindung kann das Gleichgewicht dieses Konformationsensembles verschieben und

Bindungsstellen, sogenannte kryptische Taschen, aufdecken. Kryptische Bindungsstellen entstehen nur bei der Bindung kleiner Moleküle und sind oft flach und strukturlos, so dass sie in Kristallstrukturen ohne gebundene Liganden nicht leicht zu erkennen sind. Wenn neue Bindungsstellen, einschließlich kryptischer Taschen, entdeckt werden, könnten sie potenziell für die Bindung an Liganden genutzt werden und die *druggability* verbessern. *Druggability* ist die Fähigkeit eines Proteins, kleine, arzneimittelähnliche Moleküle zu binden, und bildet die Grundlage für das Design von Medikamenten. In dieser Arbeit verwende ich physikalisch basierte, rechnerische Ansätze, um die Konformationsensembles von Bindungsstellen zu untersuchen. In allen untersuchten Systemen ist aufgrund von experimentell gelösten Strukturen bekannt, dass eine bestimmte Gruppe von Liganden Konformationsänderungen hervorrufen kann. Das Ziel ist es, den Konformationsraum, der durch die Bindung eines Liganden zugänglich wird, zu untersuchen, ohne dabei die spezifischen Ligandenstrukturen oder Details über deren Wechselwirkungen zu verwenden. Wir sind daran interessiert, das Ensemble der Taschenkonformationen zu erfassen und den jeweiligen Mechanismus der Taschenöffnung zu identifizieren. In einigen Fällen bewerten wir zusätzlich, ob die beobachtete Flexibilität ein Merkmal der Proteinfamilie oder spezifisch für das betrachtete Protein ist.

Das erste untersuchte System ist Faktor VIIa (FVIIa). FVIIa ist ein wesentlicher Bestandteil der Gerinnungskaskade und daher ein potenzieller Angriffspunkt für Medikamente für thrombotische Erkrankungen. Zusätzlich habe ich verschiedene andere Trypsin-ähnliche Serinproteasen aus derselben Proteinfamilie untersucht. Die Bindungstasche von Trypsin-ähnlichen Serinproteasen wird als S1-Tasche bezeichnet. Aufgrund einer Röntgenkristallstruktur ist bekannt, dass eine β-Faltblattstruktur in der S1-Tasche durch einen gebundenen Liganden deformiert werden kann. Ausgehend von der ungebundenen Proteinstruktur, konnte ich die Konformationsänderung mit MD-Simulationen auflösen. Ich konnte mehrere spontane Konformationsänderungen beobachten. In 7 von 22 Simulationen mit der β-Faltblattstruktur als Ausgangsstruktur ordnete sich die S1-Tasche schließlich zu einer Loopstruktur um. Diese Übergänge erfolgten spontan und wurden durch Wassermoleküle hervorgerufen, die die Wasserstoffbrückenbindungen des Rückgrats aufbrechen. Die hier untersuchte Konformationsänderung ist essentiell für die Regulierung der Substratbindung und der Katalyse. Die freie Energiebarriere dieser Konformationsänderung habe ich mit einer metadynamischen Simulation berechnet. Die geschätzte freie Energiebarriere beträgt ≈8 kJ/mol, was darauf hindeutet, dass eine Konformationsänderung in Abwesenheit eines Liganden wahrscheinlich ist, was mit unseren MD-Simulationsdaten übereinstimmt. Ich habe zusätzlich andere Serinpeptidasen untersucht, die FVIIa ähnlich sind, und stellte fest, dass die Flexibilität der S1-Tasche auch in anderen Zielproteinen wie Thrombin, Matriptase und Tryptase nachweisbar ist. Ich habe spontane Übergangsereignisse in 13 anderen Serinproteasen beobachtet, was darauf hindeutet, dass die Plastizität der S1-Tasche innerhalb der Proteinfamilie weit verbreitet ist. Unsere Ergebnisse zeigen, dass die deformierte Loopkonformation der S1-Tasche ein valides Ziel bei verschiedenen Serinproteasen ist und die Grundlage für die Entwicklung einer neuen Reihe von Inhibitoren bilden kann.

Das zweite untersuchte System ist Renin, ein Protein, das den Blutdruck im Körper re-

guliert. Hemmstoffe von Renin sind potenzielle Medikamente gegen Bluthochdruck. Die in Röntgenkristallstrukturen beobachtete Konformationsänderung von Renin ist die Öffnung eines nicht funktionalen *Flaps* in Verbindung mit einer Rotation der Trp39-Seitenkette. Ausgehend von der Proteinstruktur ohne gebundenen Liganden konnte dieser Übergang für Renin mit MD-Simulationen nicht erfasst werden. Daher verwendete ich etablierte *Biasing*-Protokolle für Renin, um das *Sampling* zu verbessern. Zunächst fügte ich organische Fragmente in die Simulationsbox hinzu, was als *mixed-solvent* MD-Simulation bezeichnet wird. Es ist bekannt, dass Fragmente in *mixed-solvent* MD-Simulationen prinzipiell Konformationsänderungen und Taschenöffnungen durch Bindung an das Protein bewirken können. Zweitens wendete ich die erweiterte *Sampling*-Methode SWISH (*Sampling Water Interfaces through Scaled Hamiltonians*) an, eine auf Replika-Austausch basierende MD-Methode, die die Wechselwirkungen zwischen Wassermolekülen und Protein verstärkt. Auf diese Weise können die Wassermoleküle mit hydrophoben Proteintaschen interagieren, was zu mehr *Sampling* von Konformationszuständen führen kann. Beide Methoden erhöhten die Flexibilität des Proteins und führten zu Fluktuationen des *Flaps*, aber die Trp39-Seitenkette blieb in ihrer ursprünglichen Konformation. Um die Energiebarriere für die Rotation der Trp39-Seitenkette zu berechnen, die das Protein davon abhält, die vollständige Konformationsänderung zu durchlaufen, habe ich metadynamische Simulationen verwendet. Die metadynamische Simulation ergab eine hohe freie Energiebarriere für die Trp39-Seitenkettenrotation ($\approx$35 kJ/mol), was die Schwierigkeiten beim *Sampling* erklärt. Die Fallstudie von Renin zeigt, dass die Bindung kleiner Moleküle mit einer hohen Energiebarriere verbunden sein und dennoch eine hohe Affinität erreicht werden kann.

Wir untersuchten auch vier Zielproteine für Arzneimittel, die bekannte kryptische Taschen haben: NPC2, p38α, Rizin und Eg5. Das NPC2-Protein ermöglicht den Cholesterintransfer und ist ein potenzielles Arzneimittelziel für die Niemann-Pick-Krankheit. Die Proteinkinase p38α spielt eine entscheidende Rolle in verschiedenen Signalwegen und wird als potenzielles Ziel für verschiedene Krankheiten wie Entzündungen und Krebs angesehen. Rizin ist ein hochgiftiges Protein aus einer Pflanze, das die Proteinsynthese hemmen kann. Eg5 ist für die Zellteilung unerlässlich und ein vielversprechendes Ziel für die Krebstherapie. Alle vier Proteine sind Teil eines Datensatzes, in dem Proteine mit bekannten kryptischen Stellen aufgeführt sind. Für alle Proteine sind Röntgenkristallstrukturen der ligandengebundenen und ungebundenen Form verfügbar. Die ligandengebundenen Strukturen weisen die offenen kryptischen Taschen auf. In allen hier untersuchten Fällen besteht die kryptische Tasche aus verschobenen, oberflächlich vergrabenen hydrophoben Seitenketten. Um die dynamischen Eigenschaften der kryptischen Taschen und den entsprechenden Mechanismus für die Taschenöffnung zu untersuchen, habe ich die gleichen drei Methoden verwendet, wie oben beschrieben: MD-Simulationen, *mixed-solvent* MD-Simulationen mit 1 M Benzolmolekülen und SWISH-Simulationen. Für alle vier Proteine konnten die Öffnung der kryptischen Taschen erfasst werden. Durch den Einsatz der verschiedenen Simulationstechniken konnten wir ein gutes Verständnis der zugrundeliegenden Bindungsmechanismen erlangen und sie im Spektrum von *induced-fit* und *conformational selection* verorten. *Induced-fit* und *conformational selection* sind zwei Extreme der möglichen Öffnungsmecha-

nismen. Beim *induced-fit* wird die Konformationsänderung ausschließlich durch die Bindung eines bestimmten Liganden erreicht, während bei der *conformational selection* der Konformationsraum bereits im ungebundenen Ensemble vorhanden ist. Bei *mixed-solvent* MD-Simulationen von NPC2 habe ich beobachtet, dass zwei Benzolmoleküle nacheinander in die kryptische Tasche eindringen. Die Bindung der Benzolmoleküle führte zur Öffnung der tiefen Tasche. Das spricht für eine bedeutende Rolle des *induced-fit* im Mechanismus der Taschenöffnung. Im Gegensatz dazu war die kryptische Tasche von p38α bereits in der MD-Simulation ohne hinzugefügte Fragmente flexibel. In *mixed-solvent* MD-Simulationen von p38α hat sich ein Benzolmolekül in die kryptische Tasche platziert, was auf ein Gleichgewicht zwischen *induced-fit* und *conformational selection* hindeutet. Im Gegensatz zu NPC2 und p38α waren die Konformationsänderungen der kryptischen Taschen von Rizin und Eg5 weitgehend unabhängig von der Anwesenheit von Fragmenten, was zeigt, dass die *conformational selection* der Hauptfaktor für den Taschenöffnungsmechanismus ist. Interessanterweise offenbarte die Benzolbindung an Rizin und Eg5 neben den bekannten kryptischen Stellen mehrere zusätzliche Bindungstaschen, die möglicherweise für weitere Studien zur Entwicklung von Arzneimitteln genutzt werden können.

Im zweiten Teil der Arbeit habe ich die Wechselwirkungen von Enzym-Substrat-Interaktionen bei zwei Wirkstoffzielen auf atomistischer Ebene untersucht: der papainähnlichen Protease (PLpro) von SARS-CoV-2 und der menschlichen Casein Kinase 1 (CK1).

PLpro spielt eine wesentliche Rolle bei der Replikation des Coronavirus, indem es virale Polyproteine prozessiert. Darüber hinaus kann PLpro von SARS-CoV-2 die angeborene Immunantwort beeinträchtigen, indem es mit höherer Aktivität ISG15 von zellulären Proteinen abspaltet. Bei verschiedenen Coronaviren (SARS-CoV und SARS-CoV-2) unterscheidet sich PLpro in der Substratpräferenz. PLpro aus SARS-CoV-2 spaltet bevorzugt ISG15 im Vergleich zu Ubiquitin, während PLpro aus SARS-CoV eine höhere Präferenz für Ubiquitin hat. Die höhere Präferenz für ISG15 von PLpro aus SARS-CoV-2 führt zu einer reduzierten angeborenen Immunantwort, was eine mögliche Erklärung für die höhere Übertragungsrate von SARS-CoV-2 im Vergleich zu früheren Coronaviren ist. Die hier durchgeführten MD-Simulationen und Analysen identifizieren die molekulare Grundlage für die experimentellen Befunde, dass PLpro bevorzugt ISG15 spaltet. Wir beobachteten, dass PLpro aus SARS-CoV-2 stärker mit ISG15 als mit Ubiquitin interagiert. Während ISG15 in drei unabhängigen MD-Simulationen von je 3,2 µs gebunden blieb, dissoziierte sich das distale Ubiquitin von PLpro in vier von sechs Durchläufen auf einer Zeitskala von Mikrosekunden. Die Simulationsdaten zeigten, dass der Dissoziationsmechanismus von Ubiquitin wasservermittelt ist. Wassermoleküle dringen zwischen die Bindungsschnittstelle ein und führen zur Ubiquitinabspaltung. Ein Vergleich der MD-Simulationen von Ubiquitin-gebundenem PLpro aus SARS-CoV und SARS-CoV-2 zeigte, dass die Punktmutation L75T im Kern der Bindungsschnittstelle der Grund für die unterschiedliche Ubiquitin-Präferenz ist, da diese den hydrophoben Cluster innerhalb der Bindungsschnittstelle schwächt. Im Gegensatz zu Ubiquitin interagiert ISG15 eng mit PLpro von SARS-CoV-2 in allen drei Simulationsreplikaten. Die pharmakologische Hemmung von PLpro in SARS-CoV-2 blockiert daher nicht nur die virale Replikation, sondern verstärkt auch gleichzeitig die antivirale

Immunantwort. Einer der ersten für PLpro entwickelten Inhibitoren ist GRL-0617. Mithilfe von MD-Simulationen fanden wir heraus, dass der Bindungsmodus von GRL-0617 für PLpro von SARS-CoV und SARS-CoV-2 derselbe ist, was darauf hindeutet, dass frühere Erkenntnisse aus der Medikamentenforschung für PLpro von SARS-CoV auf SARS-CoV-2 übertragen werden können.

CK1 steuern verschiedene wichtige zelluläre Prozesse, darunter die DNA-Reparatur. In Eizellen nach einer durch Chemotherapie verursachten DNA-Schädigung wird die CK1δ-Isoform mit der Aktivierung eines Zelltodprogramms in Verbindung gebracht, das bei Frauen zu Unfruchtbarkeit führt. Ein wichtiger Schritt in diesem Prozess ist die ratenlimitierende dritte Phosphorylierung von dem CK1-Substrat TAp63α, die TAp63α in eine aktive Konformation überführt. Um zu verstehen, warum die dritte Phosphorylierung die langsamste ist, habe ich die Enzym-Substrat-Wechselwirkungen mit MD-Simulationen auf molekularer Ebene untersucht. Wir konnten eine atypische Enzym-Substrat-Interaktion als Ursache für die langsame Kinetik dieses entscheidenden Schritts im Vergleich zu den anderen drei Phosphorylierungsereignissen ausmachen. Die Simulationsdaten enthüllten die stabilisierenden Wechselwirkungen zwischen CK1δ und TAp63α als starke Salzbrücken und enge hydrophobe Kontakte in einer für den Phospho-Transfer ungünstigen Form. Es wurden Punktmutationen eingeführt, um die mögliche funktionelle Bedeutung einzelner Aminosäuren in der Bindungsschnittstelle zu ermitteln. Diese Simulationen deuten darauf hin, dass zwei Aminosäuren von TAp63α eine wesentliche Rolle bei der Verlangsamung der Phosphorylierungskinetik spielen. Unsere Kollaborationspartner haben gezeigt, dass die Hemmung von CK1δ bei Mäusen die Eizellen intakt hält, selbst unter dem Einfluss von Chemotherapeutika. Neben der CK1δ-Isoform habe ich auch die CK1γ3-Isoform untersucht, die ein anderes Verhalten zeigt. Der dritte Phosphorylierungsschritt wird nicht verlangsamt, sondern vollständig eliminiert. Mithilfe von MD-Simulationen haben wir eine Lys-zu-Leu-Mutation als die Schlüsselmutation identifiziert, die die δ- und die γ3-Isoform voneinander unterscheidet, da das Leu257 in der γ3-Isoform die hydrophoben Wechselwirkungen mit dem TAp63α-Substrat verstärkt, was die Phosphorylierung stoppen kann. Darüber hinaus fanden unsere experimentellen Kollaborationspartner heraus, dass CK1 autophosphoryliert werden kann, was zu einer reduzierten Enzymaktivität führt. Die hier durchgeführten MD-Simulationen und Analysen geben Aufschluss über die molekulare Grundlage für die experimentellen Beobachtungen. Die Simulationsdaten zeigen, dass die autophosphorylierte Form eine größere Plastizität aufweist als die native Form. Die verringerte Aktivität wird durch die veränderte Konformation in der phosphorylierten Form verursacht.

Diese Arbeit zeigt, dass Computersimulationen eine effektive Methode sind, um Konformationsensembles von Proteinen und Enzym-Substrat-Interaktionen zu untersuchen, die zu einer rigorosen Bewertung von Wirkstoffzielen und zu molekularen Erklärungen für experimentelle Ergebnisse beitragen können. MD-Simulationen ermöglichten es uns, die Flexibilität von Bindungsstellen und Enzym-Substrat-Wechselwirkungen mit atomarer Auflösung zu untersuchen. Somit konnten wir einen Beitrag zur Bewertung kryptischer Taschen und ihrer jeweiligen Öffnungsmechanismen leisten. Wir fanden heraus, dass nicht nur Faktor VIIa eine große Flexibilität der Bindungsstelle aufweist, sondern auch ver-

schiedene weitere Zielproteine innerhalb derselben Proteinfamilie, was die Grundlage für das Design neuer Serinproteaseinhibitoren sein kann. Darüber hinaus zeigten die Simulationsergebnisse von Renin, dass die Konformationsänderungen in der Bindungsstelle mit einer hohen energetischen Barriere einhergehen können, was darauf hindeutet, dass kleine Moleküle in der Lage sind, größere Konformationsänderungen hervorzurufen und dennoch eine hochaffine Bindung zu erreichen. Die Fallstudien zu NPC2, p38α, Rizin und Eg5 zeigten, dass die Mechanismen für kryptische Taschenöffnungen systemabhängig sind. Darüber hinaus konnten wir zum Verständnis von Krankheitsmechanismen beitragen, insbesondere bei COVID-19 und Unfruchtbarkeit bei Frauen nach einer Chemotherapie. Wir haben eine molekulare Erklärung gefunden, warum PLpro von SARS-CoV-2 ISG15 besser spaltet als frühere Coronaviren, was möglicherweise ein Grund für die höhere Übertragungsrate des aktuellen Virus ist. Außerdem fanden wir einen Grund, warum die Kinetik der durch CK1-Proteine vermittelten Phosphorylierungsschritte unterschiedlich ist, indem wir die Enzym-Substrat-Interaktionen auf molekularer Ebene untersuchten. Wir zeigten, wie Simulationen genutzt werden können, um Konformationsänderungen von Proteinen und Wechselwirkungen zwischen Proteinen zu erforschen. Wir zeigten auch, wie biomolekulare Simulationen experimentelle Daten ergänzen können. Dies stimmt uns zuversichtlich, dass mit zunehmender Computerleistung und weiterentwickelten Simulationsmethoden in Zukunft auch hochenergetische Konformationsänderungen und komplexere Prozesse erfasst und verstanden werden können. Die Fortschritte bei den Berechnungsmethoden, einschließlich der physikbasierten und auf maschinellem Lernen beruhenden Ansätze, werden Forschern helfen, neue Angriffspunkte für Medikamente zu finden.

Summary

Finding new drugs is a difficult, time-consuming, and costly challenge, with only a small success rate along the drug discovery pipeline of far less than 10%. The high failure rate of drug discovery projects motivates the integration of computational tools throughout the whole drug discovery pipeline, from target identification to clinical trials. Target identification is the first step in the process. A biological target, e.g., a protein that plays a role in disease, is identified and its molecular mechanism in the disease is studied. Further, a potential binding site on the target, where therapeutic molecules can bind and modulate the target's activity, needs to be characterized. Computational tools can contribute to improving the initial molecular target elucidation and assessment.

In this thesis, I use computational, physics-based approaches to characterize binding sites of drug targets and to decipher enzyme-substrate interactions, which play a role in disease mechanisms. Molecular dynamics (MD) simulations were applied to study the dynamics of molecules in solution at high temporal and spatial resolution. The method generates time-resolved trajectories of the particles in a system of interest by integrating Newton's equations of motion numerically, starting from a set of coordinates and velocities. In MD simulations, all atoms of a chosen system, including solvent, are represented explicitly. Atomistic simulations are especially well-suited to study detailed interactions that depend on intermolecular interactions, such as hydration effects, hydrogen bonding, hydrophobic interactions, or subtle chemical differences. System properties are inferred from the trajectories, provided that the force fields, describing the interactions between the particles in the system, have a high accuracy. The bonded and non-bonded interactions are parametrized on experimental and quantum chemical data. The purpose of MD simulations can be to gain insight into the behavior of complex biological systems at molecular level, which often cannot be observed in experiments at the same resolution. With recent advances in computer hardware and simulation software, molecular systems of increasing size and simulation length can be investigated.

In the first part of the thesis, I investigated the conformational ensemble of various protein drug targets. Proteins are dynamic biomacromolecules that can have diverse and nearly isoenergetic conformational states. Ligand binding can shift the equilibrium of this conformational ensemble and can uncover binding sites, called cryptic sites. Cryptic sites only emerge upon small molecule binding and are often flat and featureless, and thus not easily recognized in crystal structures without bound ligands. If new binding sites including cryptic sites are detected, they can potentially be exploited for binding to ligands and enable a druggable target. Druggability is the ability of a protein to bind small, drug-like molecules, which is the basis for rational drug design. In this thesis, I used state-of-the-art physics-based, computational approaches to investigate the conformational ensembles of binding sites. In all studied systems, it is known from experiment that a

specific group of ligands can induce conformational changes. The aim is to sample the conformational space made accessible upon ligand binding, yet without using the specific ligand structures or details about their interactions. We are interested in sampling the pocket conformational states and identifying the respective pocket opening mechanism. For some cases, I additionally assessed whether the observed flexibility is a feature of the protein family, or specific to the protein under consideration.

The first studied system is factor VIIa (FVIIa). FVIIa is an essential part of the coagulation cascade and hence a potential drug target for thrombotic diseases. In addition, I investigated various other trypsin-like serine proteases from the same protein family. The binding pocket of trypsin-like serine proteases is called S1 pocket. An X-ray crystal structure solved by our collaborators reveals that a β-sheet structure in the S1 pocket is distorted by a bound ligand. I resolved the conformational change with MD simulations, starting from the unbound protein structure solvated in water and ions. I observed multiple spontaneous transition events. In 7 out of 22 simulations with the β-sheet as starting structure, the S1 pocket eventually rearranged into a distorted loop structure. These transitions occurred spontaneously and were mediated by water molecules probing the backbone hydrogen bonds. The conformational change studied here controls the onset of substrate binding and catalysis. Furthermore, I used metadynamics simulation, an enhanced-sampling method, to estimate the free energy barrier of this conformational change. The estimated free energy barrier is ≈ 8 kJ/mol, suggesting that interconversion is probable in the absence of a ligand, consistent with our equilibrium MD simulation. I additionally studied other serine peptidases similar to FVIIa and found that the S1 pocket flexibility is present in other drug targets including thrombin, matriptase and tryptase. I observed spontaneous transition events in 13 other serine proteases, indicating that S1 pocket plasticity is common within the protein family. Our results show that the distorted loop conformation of the S1 pocket is a valid drug target in various serine proteases and can be used as the basis for the design of a new series of inhibitors.

The second studied system is renin, which is a protein that regulates blood pressure. Inhibitors of renin are potential antihypertensive drugs. For renin, the opening of a non-functional flap together with Trp39 side chain rotation was observed in a set of X-ray crystal structures. Using MD simulations starting from the renin structure without bound ligand, the transition could not be sampled. Thus, I used established biasing protocols for renin to improve sampling. First, I tested mixed-solvent MD simulation where organic fragments are added to the simulation box. It is known that fragments in the mixed-solvent MD simulations can in principle induce conformational changes and pocket openings by binding to the protein. Secondly, I applied the enhanced sampling method SWISH (Sampling Water Interfaces through Scaled Hamiltonians), which is a replica exchange MD based method that increases the interactions between solvent molecules and protein. In this way, solvent molecules can interact with the hydrophobic protein pockets, which can lead to more sampling of conformational states. Both methods increased the protein flexibility and led to flap fluctuations, but the Trp39 side chain remained in its initial conformation. To calculate the energy barrier for the Trp39 side chain rotation, which keeps the protein from

undergoing the full conformational change, I used well-tempered metadynamics simulation. The metadynamics simulation indicated a high free energy barrier for the Trp39 side chain rotation ($\approx$35 kJ/mol), which explains the sampling difficulties. The case study of renin shows that small molecule binding can be associated with a high energy barrier and still achieve high affinity.

We also examined four protein drug targets that have established cryptic sites: NPC2, p38α, ricin and Eg5. The NPC2 protein enables cholesterol transfer and is a potential drug target for the Niemann-Pick disease. The protein kinase p38α plays a critical role in several signalling pathways and is considered as a potential target for various diseases such as inflammatory diseases and cancer. Ricin is a highly toxic protein from a plant that can inhibit protein synthesis. Eg5 is essential for cell division and is a promising target in cancer therapy. All four proteins are part of a data set, which lists proteins with known cryptic sites. For all proteins, X-ray crystal structures of the ligand-bound and unbound form are available. The ligand-bound structures exhibit the open cryptic pockets. In all cases studied here, the cryptic pocket consists of displaced shallowly buried hydrophobic side chains. To study the dynamical properties of cryptic sites and the respective mechanism for pocket opening, I used the same three methods to probe and induce the opening of cryptic pockets starting from the unbound structures as outlined above: Equilibrium MD simulations in water solvent, equilibrium mixed-solvent MD simulations with 1 M benzene molecules, and SWISH simulation. For all four proteins, the opening of the cryptic pockets was captured. By analyzing the various simulations, we were able to acquire a comprehensive understanding of the underlying binding mechanisms. The observed mechanisms range from induced-fit to conformational selection. Induced-fit and conformational selection are two extremes of possible opening mechanisms. In induced-fit, the conformational change is solely triggered by binding of a specific ligand, whereas in the conformational selection the conformational space is already present in the unbound ensemble. In mixed-solvent MD simulations of NPC2, I observed that two benzene molecules sequentially entered into the cryptic pocket. Binding of the benzene molecules resulted in opening of the deep pocket, which points towards a significant role of induced-fit in the pocket opening mechanism. The cryptic pocket of p38α, in contrast, was already flexible in the MD simulation without added fragments. In mixed-solvent MD simulations of p38α, a benzene molecule inserted into the cryptic pocket, suggesting a balance of induced-fit and conformational selection. Contrary to NPC2 and p38α, the conformational changes of the cryptic pockets of ricin and Eg5 were largely independent of the presence of probe molecules, demonstrating that conformational selection is the main factor in the pocket opening mechanism. Interestingly, benzene binding to ricin and Eg5 revealed multiple additional binding pockets besides the known cryptic sites, which can potentially be used for further drug design studies.

In the second part of the thesis, I obtain atomistic insight into the interactions of enzyme-substrate interactions in two drug targets: the papain-like protease (PLpro) of SARS-CoV-2 and the human casein kinase 1.

PLpro plays an essential role in coronavirus replication by processing viral polyproteins. In addition, PLpro can interfere with the innate immune response by cleaving the substrate

ISG15. PLpro substrate preference differs for different corona viruses (SARS-CoV and SARS-CoV-2). PLpro from SARS-CoV-2 preferentially cleaves ISG15 compared with ubiquitin, whereas PLpro from SARS-CoV has a higher preference for ubiquitin. The higher preference for ISG15 of PLpro from SARS-CoV-2 results in a reduced innate immune response, which is a potential explanation for the higher transmission rate of SARS-CoV-2 compared with earlier corona viruses. The MD simulations and analyses conducted here identified the molecular mechanism for the experimental findings that PLpro preferentially cleaves ISG15. I observed that PLpro from SARS-CoV-2 interacts more strongly with ISG15 compared with ubiquitin. While ISG15 remained bound in three independent MD runs of 3.2 μs, the distal ubiquitin of di-ubiquitin dissociated from PLpro in four out of six runs on a microsecond timescale. The simulation data revealed that the dissociation mechanism of ubiquitin is water-mediated. Water molecules enter in between the binding interface leading to ubiquitin separation. Comparison of ubiquitin-bound PLpro from SARS-CoV and SARS-CoV-2 showed that the point mutation L75T in the core of the binding interface is the reason for the difference in ubiquitin preference, as it weakens the hydrophobic cluster within the binding interface. In contrast to ubiquitin, ISG15 tightly interacts with PLpro from SARS-CoV-2 in all three simulation replicas. Hence, pharmacological inhibition of PLpro in SARS-CoV-2 not only blocks viral replication but also simultaneously boosts the antiviral immune response. One of the first inhibitors designed for PLpro is GRL-0617. Using MD simulations, I found that the binding mode of GRL-0617 is the same for PLpro from both SARS-CoV and SARS-CoV-2, suggesting that findings from drug design efforts on PLpro from the earlier SARS-CoV can be transferred to targeting SARS-CoV-2.

Casein kinases 1 (CK1) control various essential cellular pathways, including DNA repair. In oocytes after chemotherapy-induced DNA damage, the CK1δ isoform is associated with activating a programmed cell-death that results in infertility in women. A key step in this process is the rate-limiting third phosphorylation of the CK1-substrate TAp63α, which converts TAp63α into an active conformation. To understand, why the third phosphorylation is the slowest, I investigated the enzyme-substrate interactions at a molecular level using MD simulations. We could pinpoint an atypical enzyme-substrate interaction as the cause of the slow kinetics of this decisive step, compared to the other three phosphorylation events. The simulation data revealed the stabilizing interactions between CK1δ and TAp63α as strong salt bridges and tight hydrophobic contacts in a form unfavorable for phospho-transfer. Point mutations were introduced to identify the possible functional relevance of single amino acids in the binding cleft, suggesting that two key residues of TAp63α play a significant role in decelerating the phosphorylation kinetics. Our collaborators showed that inhibiting CK1δ in mice keeps the oocytes intact, even under the influence of chemotherapeutic agents. In addition to the CK1δ isoform, we investigated the CK1γ3 isoform, which shows a different behavior. Here, the third phosphorylation step is not decelerated, but fully eliminated. Using MD simulations, I identified a Lys to Leu mutation as the key mutation distinguishing the δ and γ3 isoform, as Leu257 in the γ3 isoform strengthens the hydrophobic interactions to the TAp63α substrate, which potentially stalls phosphorylation. In addition, our experimental collaborators found that

CK1 can be autophosphorylated, leading to reduced enzyme activity. Our simulations shed light on the molecular basis for the experimental observations. They revealed that the autophosphorylated form displayed greater plasticity than the native form. The decreased activity is caused by the disruption of the integrity of the substrate binding site in the phosphorylated form.

This thesis shows that computational simulations are a powerful tool to study conformational ensembles of proteins and enzyme-substrate interactions providing rigorous drug target assessment and molecular explanations to experimental findings. MD simulations allowed us to investigate binding site flexibility and enzyme-substrate interaction in full atomic detail. I could contribute to the assessment of cryptic pockets and their respective pocket opening mechanism. I found that the known binding site flexibility of FVIIa can be extended to various drug targets within the same protein family, which might aid rational drug design of new serine protease inhibitors. Furthermore, the simulation results of renin showed that the conformational changes in the binding site can come with a high energetic barrier indicating that small molecules are able to induce major conformational changes, and still achieve high-affinity binding. The case studies of NPC2, p38α, ricin, and Eg5 show that the mechanisms for cryptic pocket openings are system dependent. In addition, we could contribute to the understanding of disease mechanisms, in particular COVID-19 and infertility in women after chemotherapy. I found a molecular explanation to why PLpro from SARS-CoV-2 cleaves ISG15 better compared to earlier corona viruses, which is potentially a reason for the higher transmission rate of the current virus. Further, we rationalized why the kinetics of the phosphorylation steps mediated by CK1 proteins are different by observing the enzyme-substrate interactions at a molecular level. We demonstrated how simulations can be used to study protein conformational changes and interactions between proteins. We also showed how biomolecular simulations can complement experimental data. This makes us confident that also high-energy conformational changes and more complex processes can be captured and understood in the future with increasing computer power and enhanced sampling methods. The advances in computational methods, including physics-based and machine learning based approaches, will help researchers to identify new druggable targets.

Introduction

The fundamental understanding of disease mechanisms and the identification of druggable biomolecular targets are the basis for rational drug discovery research. In this thesis, I used physics-based simulation approaches to contribute both to the druggability assessment of specific drug targets and to the understanding of disease mechanisms, specifically COVID-19 and infertility in women after chemotherapy.

1.1   Cryptic pockets in protein drug targets

Drug discovery is a challenging, complex, and multidisciplinary effort. Many approved drugs are small molecules and function by binding to various specific drug targets. Drug targets can be proteins, such as enzymes, ion channels, and G-protein-coupled receptors. Binding of small molecules to proteins can alter their functions and hence trigger the desired biological response, affecting the symptom or the disease that the drug was intended to treat. Proteins, acting as receptors, bind small molecules, called ligands or compounds, with high affinity and specificity. The ability of a protein to bind to other molecules is determined by the physicochemical properties of the residues forming its binding pocket and the shape of the pocket. However, a wide range of proteins does not have binding pockets on the protein surface. Up to 85% of the human proteome lacks easy-to-find pockets [Neklesa et al., 2017] where therapeutics in form of, e.g., small molecules can bind. The vast majority of the proteome is perceived as "undruggable". Identifying new cryptic sites can provide druggable targets, which were previously assumed to be undruggable. Only a small proportion of the proteome has known "hidden" binding sites that only become apparent upon ligand binding. These cryptic pockets are not detectable in the ligand-free structures. Ligand binding can induce pocket opening by shifting the equilibrium of the protein conformational ensemble (Fig. 1.1). Cryptic pockets are not necessarily "functional" pockets, but despite their distance from the active site, several cryptic sites have been shown to modulate the protein's activity [Lu et al., 2018].

There exists a continuous range of pocket forming mechanisms between the extremes "induced-fit" and "conformational selection" [Beglov et al., 2018]. In the induced fit mechanism, the protein pocket can open up only with and through the insertion of a ligand. In the conformational selection mechanism, the conformation of the protein changes independently of the ligand and binding can occur when the ligand encounters a conformation of the protein in which the pocket is already fully open.

Some of the first discovered cryptic pockets were found serendipitously thanks to buffer molecules binding into the cavity [Horn and Shoichet, 2004]. Cryptic pockets are usually
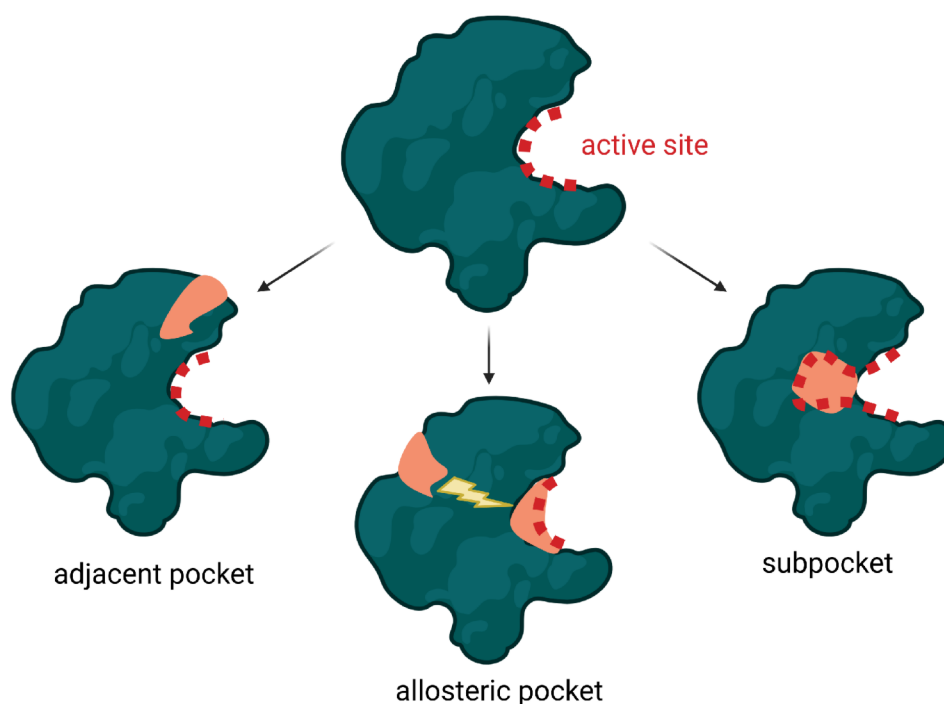
Figure 1.1: **Cryptic pockets.** Cryptic pockets can be present in the form of adjacent pockets (left), allosteric pockets (middle), and subpockets within the active site (right). Created with BioRender.com.

investigated using screening libraries or site directed tethering [Hardy and Wells, 2004]. However, these approaches require significant effort that was often not worthwhile. Therefore, computational methods for identifying and studying cryptic pockets are of growing interest as available computer power increases.

In this thesis, I compared molecular dynamics(MD)-based detection methods in terms of efficiency for finding cryptic pockets and uncovering plasticity of binding pockets of pharmacologically interesting targets. I investigated the plasticity of the active sites of factor VIIa and renin and of proteins within the same protein family. Additionally, I selected four proteins of the dataset CryptoSite to investigate cryptic pockets: NPC intracellular cholesterol transporter 2 (NPC2), mitogen-activated protein kinase 14 (p38α), ricin, and Eg5. In the following, I introduce these systems in detail.

1.1.1 Factor VIIa

The coagulation factor VIIa (FVIIa) is a serine protease essential for blood clotting. The complex of FVIIa with the membrane-bound tissue factor (TF) initiates the coagulation cascade, resulting in fibrin formation. Therefore, FVIIa is a promising target for treating thrombotic diseases and providing anticoagulant alternatives. In the past, considerable efforts have been made to design selective inhibitors of the FVIIa-TF complex. The development of clinical-stage FVIIa-TF-based anticoagulants was not successful, primarily due to the lack of oral bioavailability [Zbinden et al., 2005, Shirk and Vlasuk, 2007, Schreuder
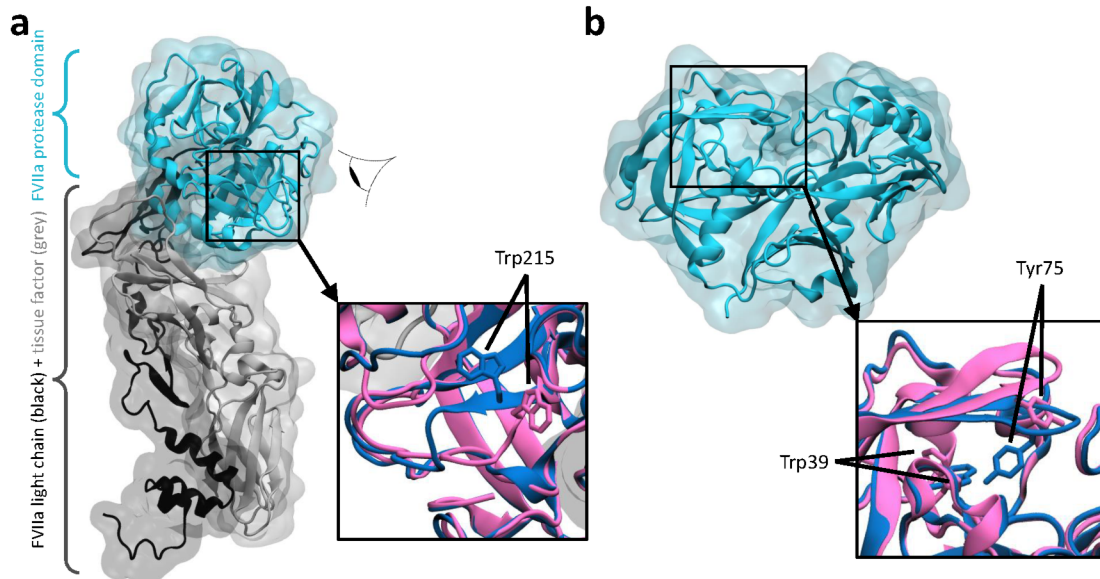
Figure 1.2: **Remodeled active sites of FVIIa and renin. a,** X-ray crystal structure of FVIIa [Sanofi in-house structure] with zoom into its active site. Zoom inset displays β-sheet conformation (blue) [Sanofi in-house structure] and loop-like conformation at Trp215 (pink) [Sanofi in-house structure]. **b,** X-ray crystal structure of renin [PDB entry: 1rne] with zoom into its active site. Zoom inset shows closed flap (blue) [PDB entry: 1rne] and open flap conformation (pink) [PDB entry: 3g72].
Protein is shown in cartoon, important residues for conformational change are displayed with atomic detail. Ligand structures are not shown.

and Matter, 2020].

FVIIa is a multidomain enzyme with an N-terminal g-carboxyglutamic acid-containing domain, two epidermal growth factor-like domains, and a C-terminal serine protease domain. The enzyme consists of a light and a heavy chain linked by a disulfide bond. The characteristic feature of trypsin-like serine proteases, including FVIIa, is a deep S1 binding pocket with an acidic Asp189 at the bottom [Banner et al., 1996, Eigenbrot et al., 2001]. Here and in Chapter 4, I consistently use chymotrypsin numbering.

In contrast to most known FVIIa inhibitors, an oxazole-derivative inhibitor of FVIIa binds to a remodeled S1 pocket and distorts the β-sheet structure at Trp215 (Fig. 1.2a) [Tesmer et al., 2022].

1.1.2 Renin

Renin is an aspartyl protease that is part of the renin-angiotensin-aldosterone system (RAAS). RAAS regulates the body's fluid and electrolyte balance.

RAAS is a series of enzymatic cleavages designed to regulate blood pressure. Renin is secreted in the cells of the juxtaglomerular apparatus in vertebrate kidneys, when the juxtaglomerular apparatus senses a drop in blood pressure. By converting angiotensinogen to angiotensin 1, renin initiates the enzyme series to increase blood pressure. Angiotensin 1 is cleaved to angiotensin 2 by the angiotensin converting enzyme, produced primarily in the endothelial cells in the lungs. Active angiotensin 2 leads to vasoconstriction in the blood

vessels and to increased production of a specific steroid hormone, which mainly causes sodium and water retention in the kidneys. In this way, angiotensin 2 upregulates blood pressure both by contracting the blood vessels and by increasing the intravascular volume [Fountain and Lappin, 2021].

RAAS can be blocked pharmacologically at several points to treat hypertension, including inhibiting renin and the angiotensin converting enzyme. Renin inhibitors target its active site and can induce conformational changes, including lifting its characteristic β hairpin structure known as flap (Fig. 1.2b).

1.1.3 Database Cryptosite: NPC2, p38α, ricin, Eg5

The best evidence for a cryptic pocket is an X-ray crystal or cryo-EM protein structure determined with a ligand binding to an unusual binding pocket. Cimermancic et al. [2016] systematically searched for cryptic binding pockets in all available protein structures in the Protein Data Bank (PDB) and created the dataset CryptoSite. CryptoSite lists a set of unbound and ligand-bound pairs with cryptic sites, consisting of 92 apo-holo pairs. Beglov et al. [2018] further expanded the dataset, currently containing 4950 structures, and investigated local flexibility at cryptic sites.

Cimermancic et al. [2016] generated the dataset CryptoSite in a systematic manner. They collected protein-ligand complexes and corresponding unbound structures from the PDB. The binding site was defined as the residues with at least one atom within 5 Å from any ligand atom in the bound conformation. After removal of redundant protein occurrences by applying a sequence identity threshold of 40%, they evaluated each binding site using the pocket scores Fpocket and ConCavity. They added all protein structures to the dataset CryptoSite whose binding sites had a bad score in the unbound and a good score in the ligand-bound conformation.

We selected four pharmacologically relevant proteins from dataset CryptoSite for investigation additional to factor VIIa and renin: p38α, NPC2, Eg5, and ricin. With this dataset, I tested current computational state-of-the art methods to decipher cryptic sites. I compared conventional MD simulations (see Chapter 3.1), MD simulations in mixed-solvent solution, and a special kind of replica exchange simulations specifically designed for uncovering cryptic sites, called SWISH (Sampling water interfaces through scaled Hamiltonians). The latter is a Hamiltonian replica exchange simulation scheme, varying the protein-solvent interaction in each replica (see Chapter 3.2). It was developed in the group of Gervasio [Oleinikovas et al., 2016] with the idea to "hydrate" hidden cryptic sites leading to openings. Similar to Comitani and Gervasio [2018], here I sampled the pocket openings of p38α and NPC2, and applied the same methods to two unexplored proteins Eg5 and ricin. In the following, I will shortly introduce the selected protein systems.

p38α is a mitogen-activated protein kinase that is activated by external stimuli such as cytokines, radiation, heat, and osmotic shock. The kinase is involved in signaling cascades and is important for cell differentiation, cell growth, and apoptosis. Modulating the activity of p38α is relevant for the treatment of a variety of inflammatory and autoimmune diseases [Hui et al., 2007]. While most ligand-bound structures in the PDB database have inhibitors
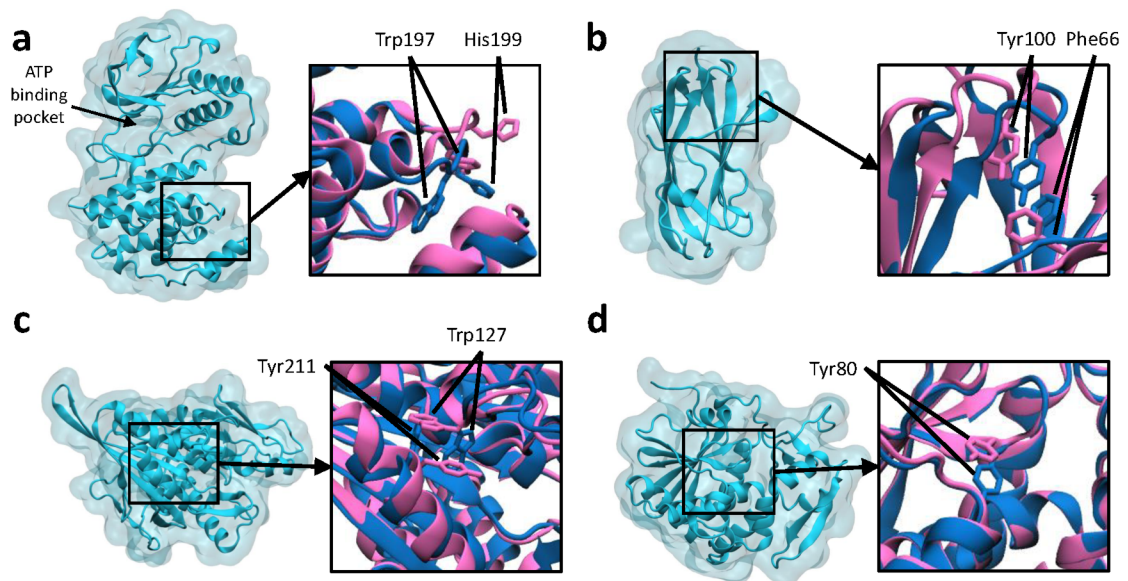
Figure 1.3: **Cryptic pockets in p38α, NPC2, Eg5, and ricin. a,** X-ray crystal structure of p38α [PDB entry: 4e5b] with zoom into its lipid-binding site. Zoom inset displays apo (blue) [PDB entry: 4e5b] and ligand-bound conformation (pink) [PDB entry: 3hl7]. **b,** X-ray crystal structure of NPC2 [PDB entry: 1nep] with zoom into its orthosteric binding site. Zoom inset shows apo (blue) [PDB entry: 1nep] and ligand-bound conformation (pink) [PDB entry: 2hka]. **c,** X-ray crystal structure of Eg5 [PDB entry: 3hqd] with zoom into the binding site 12 Å away from its orthosteric binding site. Zoom inset displays apo (blue) [PDB entry: 3hqd] and ligand-bound conformation (pink) [PDB entry: 1q0b]. **d,** X-ray crystal structure of ricin [PDB entry: 1rtc] with zoom into its orthosteric binding site. Zoom inset shows apo (blue) [PDB entry: 1rtc] and ligand-bound conformation (pink) [PDB entry: 1br6].

Protein is shown in cartoon, important residues for conformational change are displayed with atomic detail. Ligand structures are not shown.

bound to the orthosteric ATP binding site, the X-ray crystal structure from Xing et al. [2009] shows in addition a diaryl pyrazole compound bound to the lipid-binding site of the C-terminal cap [PDB entry: 3hl7]. The compound rearranges the side chain of Trp197 and displaces His199 with respect to the apo conformation [PDB entry: 4e5b [Tzarum et al., 2012]] (Fig. 1.3a).

NPC2 is a soluble lysosomal glycoprotein that binds and transfers sterols and plays a critical role in the clearance of cholesterol from the lysosomal compartment in conjunction with NPC1. NPC2 transports unesterified cholesterol to the cholesterol binding site of the membrane bound NPC1. Mutations in NPC2 cause a lethal neurovisceral disorder, called the Niemann-Pick type C2 disease, leading to the accumulation of cholesterol in lysosomes [Infante et al., 2008]. The X-ray crystal structure of the cholesterol-3-O-sulfate-bound conformation of NPC2 [PDB entry: 2hka [Xu et al., 2007]] shows the ligand wedging between two beta-sheets, resulting in a wider pocket compared to the apo conformation [PDB entry: 1nep [Friedland et al., 2003]]. The ligand expands the pocket and leads to side-chain flipping of Phe66 (Fig. 1.3b).

Eg5 is a homotetrameric molecular motor protein that is crucial in cell division. The protein from the kinesin family is essential in mitosis and is involved in the assembly and maintenance of the mitotic spindle [Sawin et al., 1992, Wojcik et al., 2013]. Loss of Eg5 leads to defective centrosome separation [Castillo and Justice, 2007]. During interphase, Eg5 is associated with ribosomes and increases translation efficiency [Bartoli et al., 2011]. In the ligand-bound X-ray crystal structure from the CryptoSite data set [PDB entry: 1q0b [Yan et al., 2004]] two compounds are present. An ADP molecule is bound to the catalytic center and a monastrol is bound to a cryptic pocket 12 Å away from the active site. The monastrol displaces Trp127 and induces a second binding pocket onto the protein surface, in contrast to the structure without an inhibitor [PDB entry: 3hqd, [Parke et al., 2010]] (Fig. 1.3c). Binding to the cryptic pocket leads to inhibition of Eg5, making the cryptic pocket a promising target for cancer treatment.

Ricin is an extremely toxic protein from the seeds of the castor oil plant, *Ricinus communis.* It is composed of two polypeptide chains, the cell surface binding chain B, which facilitates cell uptake, and the enzymatically active chain A. Chain A is the carrier of the toxic effect and inactivates the 60S subunit of eukaryotic ribosomes, resulting in inhibiting protein biosynthesis [Olsnes and Kozlov, 2001]. Due to its toxicity, ricin is listed as a weapon of war in the German War Weapons Control Act [WWL]. Ricin's active site recognizes a specific adenine base on ribosomal ribonucleic acid (rRNA). Pteroic acid ligands can function as small inhibitors by binding to its active site [PDB entry: 1br6 [Yan et al., 1997]]. The pterin ring displaces the side chain of Tyr80, with respect to the apo structure [PDB entry: 1rtc [Mlsna et al., 1993]] (Fig. 1.3d).

1.2   Enzyme-substrate interaction in protein drug targets

Enzymes play a key role in metabolism. Without them, biochemical reactions would occur at rates too slow to sustain life. Most enzymes are proteins, with the exception of ribozymes, which consist of RNA. An enzyme is a biochemical catalyst that accelerates biochemical
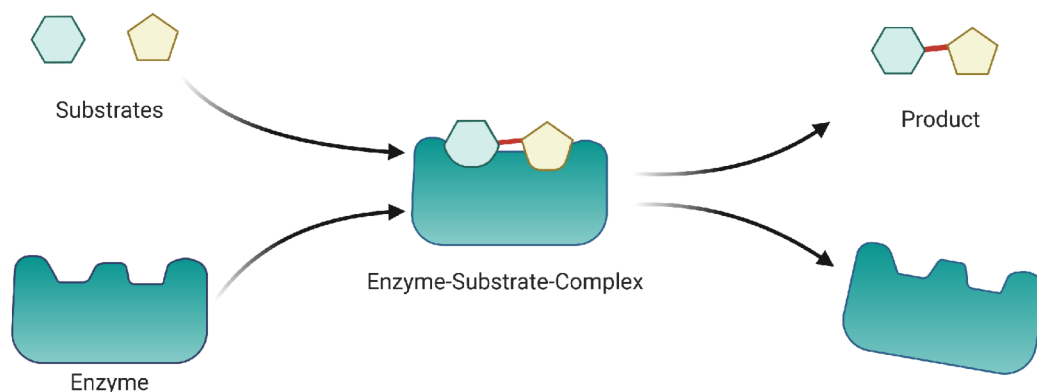
Figure 1.4: **Sketch of enzymatic mechanism.** Substrates bind to the enzyme, forming the enzyme-substrate complex, which then releases the newly built product. Upon binding, the enzyme and substrates can undergo dynamic conformational changes known as induced-fit. Created with BioRender.com.

reactions. It is not consumed in chemical reactions, nor does it modify the equilibrium of a reaction. By lowering the activation energy of reactions, it increases the reaction rate. In some cases, enzymes can make a reaction even millions of times faster [Bugg, 2012]. The enzyme always returns to its original state after the reaction is complete, and remains unchanged.

Enzymes convert substrates into one or more products through active site binding to the substrate and formation of an enzyme-substrate complex. This complex lowers the activation energy and speeds up the reaction by providing certain chemical groups or ions essential for the reaction process or by bringing substrates together in an optimal orientation. After catalyzing the reaction, the enzyme releases its product(s), as schematically shown in Fig. 1.4.

Enzymes are highly substrate-specific. The active site consists of a unique combination of amino acid residues. The positions and properties of these residues create the very specific chemical environment of the active site. Only substrates matching this site can be converted. Therefore, an enzyme only converts certain substrates to products. The enzyme's activity can be modified by other molecules. Inhibitors can decrease and activators can increase enzyme activity. Several therapeutic drugs and poisons are enzyme inhibitors. Enzymes only function well in a specific temperature and pH range.

Characterizing the interaction of an enzyme with its physiological substrates is essential for understanding disease mechanisms. In this thesis, I investigated enzyme-substrate interaction of two highly relevant drug targets: the papain-like protease of SARS-CoV-2 with interferon-stimulated gene 15 (ISG15) and ubiquitin (Ub), as well as the human CK1δ with TAp63α.
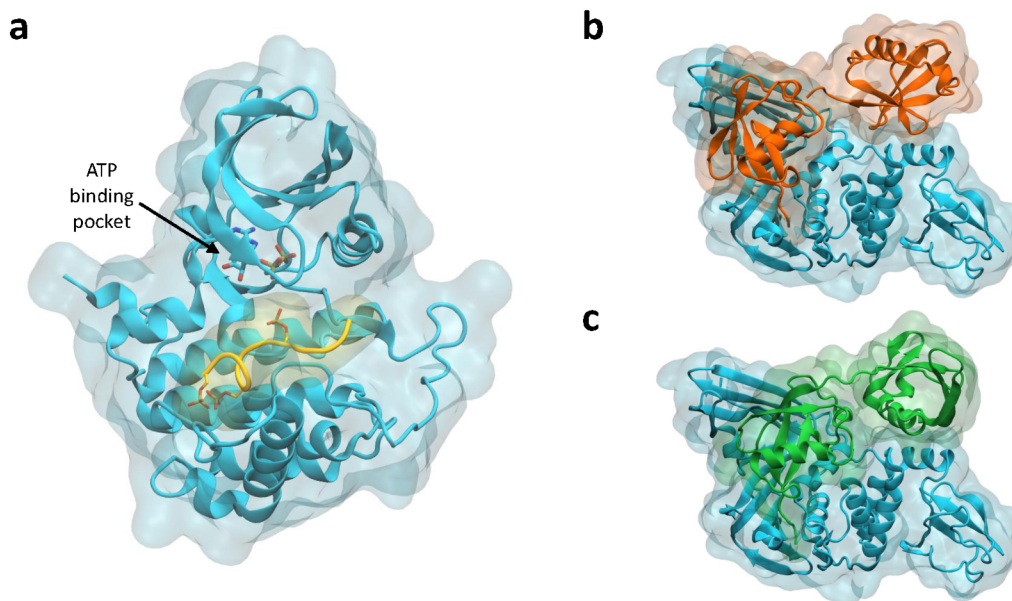
Figure 1.5: **X-ray crystal structures of CK1 and PLpro with its substrates. a,** X-ray crystal structure of CK1δ [PDB entry: 6ru8]. CK1δ (cyan) shown in cartoon with bound ADP molecule with atomic detail. Bound TAp63α peptide (yellow) shown in ribbon with three phosphorylated serines with atomic detail. **b,** Modeled structure of PLpro from SARS-CoV-2 (cyan) with bound Ub (orange) shown in cartoon (for modeling details see model 2 in Chapter 7.2). **c,** X-ray crystal structure of PLpro from SARS-CoV-2 (cyan) with bound ISG15 (green) shown in cartoon [PDB entry: 6yva, modeled missing side chains (see Chapter 7.2)].

1.2.1 Papain-like protease

Papain-like proteases (PLpro) are cysteine protease enzymes, which are present in all domains of life. PLpro in coronavirus functions as a monomer and is, in conjunction with the 3C-like protease, responsible for the virus replication by processing the viral polyproteins into functional, mature proteins. Therefore, both proteases are attractive antiviral drug targets.

In addition to processing of viral polyproteins, PLpro is involved in the removal of two cellular substrates from host cell proteins, namely ISG15 and Ub (Fig. 1.5b,c), termed deISGylation and deubiquitination, respectively. PLpro from SARS-CoV-2 preferentially removes ISG15 compared with Ub, which suppresses the host innate immune response [Shin et al., 2020]. In Chapter 7, I investigated how the altered binding interface between PLpro and Ub in the coronavirus SARS-CoV-2 weakened its activity using MD simulations.

1.2.2 Casein kinase 1

The casein kinase 1 (CK1) family represents a unique subgroup within the serine/threonine-selective enzymes. They comprise seven isoforms in humans, CK1α, β (αL), γ1-3, δ and ε, and regulate various essential cellular pathways. The multifunctional kinases play an important role in Wnt signaling [Bernatik et al., 2011, Bryja et al., 2007, Casagolda et al., 2010, Greer and Rubin, 2011, Peters et al., 1999, del Valle-Pérez et al., 2011], DNA repair [Dhillon and Hoekstra, 1994, Greer et al., 2017], circadian rhythms [Eng and Virshup, 2017, Fan et al., 2009, Meng et al., 2008, Yang et al., 2017] and mitotic checkpoint signaling

[Johnson et al., 2013]. The monomeric enzymes CK1s are largely cofactor independent and ubiquitous throughout cells and tissues [Cheong and Virshup, 2011, Knippschild et al., 2014]. CK1 kinases consist of a conserved catalytic domain (53%-98% sequence identity) at the N-terminus and a very diverse regulatory domain responsible for determining substrate specificity at the C-terminus [Cheong and Virshup, 2011].

CK1 enzymes can phosphorylate a broad range of substrates, including TAp63α, which controls oocyte quality (Fig. 1.5b). CK1δ phosphorylates TAp63α four times, resulting in apoptosis of the damaged oocyte. The third phosphorylation step is the slowest and decisive step. Using MD simulations together with NMR mutational studies, we identified the enzyme-substrate interaction sites responsible for decreasing the kinetics of the third phosphorylation step (see Chapter 8).

Autophosphorylation of the carboxy-terminal residues inhibits the kinase activity [Cegielska et al., 1998, Gietzen and Virshup, 1999, Graves and Roach, 1995, Hoekstra et al., 1994]. Additional to the carboxy-terminal autophosphorylation, autophosphorylation at the αG segment next to the catalytic site can also take place and decrease activity, as identified by our experimental collaborators [Cullati et al., 2022]. I studied this process using MD simulations at a molecular level and observed that phosphorylation destabilizes the N-terminal part of the αG helix (see Chapter 8).

Chapter 2

Aims of the thesis

The primary aim of this thesis is to contribute to identifying altered protein conformations with respect to known structures using simulation techniques based on molecular dynamics (MD). The secondary aim is to employ the same simulation tool to also study the interactions of enzymes with their substrates.

In the first part of the thesis (Chapter 4-6), the predictive power of computational methods regarding binding site plasticity is assessed. In all studied systems, it is known from X-ray crystal structures that a specific group of ligands targeting the binding site can induce conformational changes. Our aim was to sample the conformational space made accessible upon binding of the ligands, yet without using the specific ligand structures or details about their interactions. In this way, I can identify the respective pocket opening mechanism and can assess whether the observed flexibility is a feature of the protein family, or specific to the protein under consideration.

Chapter 4 focuses on the S1 pocket plasticity of factor VIIa. Unbiased MD simulations can sample the S1 pocket flexibility, and this finding can be extended to a large number of other serine proteases.

Chapter 5 deals with the known conformational change of renin and related proteins: the flap opening. I used unbiased and biased MD simulations to investigate this conformational change and evaluate the free energy barrier for the full pocket opening.

In Chapter 6, I compared the sampling efficiency of binding pocket openings in four drug targets from the CryptoSite dataset (NPC2, p38α, ricin, and Eg5) and placed their binding mechanism in the spectrum of conformational selection and induced-fit.

The second part (Chapter 7-8) deals with enzyme-substrate interactions of the papain-like protease of SARS-CoV-2 and the human casein kinase 1. MD simulations allowed us to investigate enzyme-substrate interaction in atomic detail and to contribute to the understanding of disease mechanisms, in particular of COVID-19 and infertility in women after chemotherapy.

SARS-CoV-2 is responsible for the coronavirus disease 2019 (COVID-19) pandemic. The virus depends on its papain-like protease processing the viral polyproteins. In Chapter 7, I investigated the difference in substrate interaction of PLpro between the coronaviruses SARS-CoV and SARS-CoV-2, which is one reason for the higher risk for humans arising from the latter.

CK1s regulate a variety of important cellular pathways, including DNA repair. Many chemotherapeutic agents induce DNA damage, which does not impact most healthy cells, except for oocytes. Upon DNA damage, CK1s initiate programmed oocyte death by phos-

phorylating TAp63α four times, leading to infertility in women. The third phosphorylation step is the decisive, but slowest step. In this step, TAp63α is converted into the active conformation. In Chapter 8, I elucidated the structural mechanism for the difference in the kinetic behavior based on an unusual enzyme-substrate interaction. Additionally, I studied the structural effects of autophosphorylation of the kinase domain.

Chapter 3

Computational methods

Biomolecular simulations allow one to imitate the time evolution of complex biological systems. They can reproduce experimentally observable quantities and provide insight into the mechanistic details of processes at very high spatial and temporal resolution. This can help practitioners better understand the process of interest and assist with the formulation of new hypotheses.

In this Chapter, I introduce three different simulation types applied in this thesis. First, the molecular dynamics (MD) technique is described. MD is a broadly applied and successful tool, but it is currently limited to the investigation of events that take place on microsecond time scales, unless excessive computational resources are used. Hence, MD is ill-suited to capture many interesting phenomena like cryptic pocket openings, and large conformational changes in general. To address the limited sampling time in MD, many algorithms have been developed in the past to enhance sampling. Enhanced sampling approaches build upon conventional MD. Two of them, Replica Exchange Method and Metadynamics, are introduced in the following.

3.1 Molecular dynamics simulations

MD simulation is an *in silico* method to calculate time-dependent behavior of molecular systems by numerically solving the classical equations of motion of atoms interacting via a given potential. This potential, also called force field, describes the interactions between covalent and non-covalent bonded atoms. In this section, the most fundamental aspects of MD simulations are discussed (following Chapter 4 of Frenkel and Smit [2002]).

3.1.1 Molecular force field

In classical MD simulations, force fields approximate the quantum mechanical (QM) energy surface [Vanommeslaeghe et al., 2014]. This approximation allows assessing microsecond timescales [Monticelli and Tieleman, 2013]. In contrast, QM calculations, with an accurate representation of electrons and chemical bonds, are limited to the investigation of small systems, or small reactive regions of larger systems [Senn and Thiel, 2009] with short timescales, usually of a few tens of picoseconds [Hug, 2013].

For classical force fields, the Born-Oppenheimer approximation [Born and Oppenheimer, 1927] allows for the description of atomic interactions employing the potential energy as a function of nuclear coordinates without explicit modeling of electrons. Classical force fields treat atoms as mass points in space and are parameterized based on experimental data
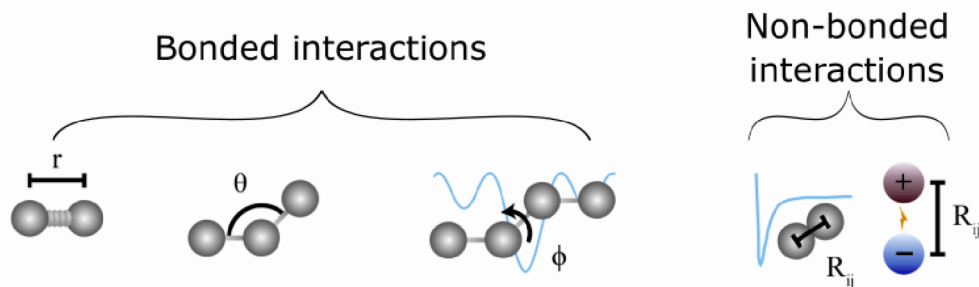
Figure 3.1: **Schematic representation of an MD force field.** The forces for molecular simulations include bonded (left) and non-bonded interactions (right). Bonded interactions consist of bonds (r), angles (θ) and dihedrals (ϕ). Bonds and angles are modeled as simple springs, and dihedrals are modeled using a sinosoidal function. Non-bonded interactions are approximated with the Lennard-Jones potential and Coulomb's law is used to model electrostatic interactions. R_{ij} denotes the distance between the two particles. Adapted with permission from Durrant and McCammon [2011] under a CC BY 2.0 license.

or QM calculations. Further, most classical force fields rely on certain assumptions, e.g., the concepts of additivity and transferability. Energy potentials in force fields are additive, and force field parameters are transferable, which assumes that the same parameter values can be applied to molecules with similar chemical structures.

Since the potential energy is additive, the force field is a sum of potential energy terms consisting of bonded and non-bonded interactions between the N particles in the system (see Fig. 3.1),

$$U(\vec{r}^N) = U_{\text{bonded}}(\vec{r}^N) + U_{\text{non-bonded}}(\vec{r}^N) , \qquad (3.1)$$

which together with the kinetic energy as a function of momenta, $K(\vec{p}^N)$, build the Hamiltonian of the system, $H(\vec{r}^N, \vec{p}^N)$.

In general, the bonded interactions are composed of energy terms for covalent bonds, angles, proper and improper dihedrals. The four atoms in an improper dihedral angle are not connected by three sequential covalent bonds. The energy terms describe oscillations around the equilibrium bond length, bond angle, and torsional rotation of four atoms about a central bond, respectively,

$$U_{\text{bonded}}(\vec{r}^N) = U_{\text{bonds}}(\vec{r}^N) + U_{\text{angles}}(\vec{r}^N) + U_{\text{dihedrals}}(\vec{r}^N) + U_{\text{improper}}(\vec{r}^N) . \qquad (3.2)$$

The first three terms express bond stretching, angle bending, and torsional rotations. The fourth term is added to ensure the preservation of planarity of molecules with a flat geometry, like aromatic rings. Bond stretching and angle bending are approximated by harmonic potentials, which is reasonable for small deviations from the equilibrium value. The potential energy for proper dihedrals at equilibrium is characterized by cosine functions with multiplicity n and phase factor ϕ_0, which depend on the bond geometry. Improper dihedral angles are restrained around their equilibrium values via harmonic potentials,

$$U_{\text{bonded}} = \frac{1}{2} \sum_{\text{bonds}} k_r (r - r_{\text{eq}})^2 + \frac{1}{2} \sum_{\text{angles}} k_\Theta (\Theta - \Theta_{\text{eq}})^2 +$$
$$\frac{1}{2} \sum_{\text{dihedrals}} k_\phi (1 + \cos(n\phi - \phi_0)) + \frac{1}{2} \sum_{\text{improper}} k_\Psi (\Psi - \Psi_{\text{eq}})^2 .$$

$$(3.3)$$

The different force constants k, the equilibrium values r_{eq} (distance), Θ_{eq} (angle), and Ψ_{eq} (torsion angle), the multiplicity n, and the phase factor ϕ_0 are free parameters that have to be tuned to fit experimental data or QM calculations for each molecule and force field.

The non-bonded potentials arise from electrostatic (Coulomb) and van der Waals (vdW) interactions,

$$U_{\text{non−bonded}}(\vec{r}^N) = U_{\text{Coulomb}}(\vec{r}^N) + U_{\text{vdW}}(\vec{r}^N) . \qquad (3.4)$$

The charge distribution of molecules is discretized in classical force fields by placing partial charges on atomic nuclei. For the determination of the partial charges, usually the restrained electrostatic potential (RESP) [Bayly et al., 1993] fitting is applied. In this method, point charges are placed into the molecule in a way that this configuration reproduces approximately the electrostatic potential calculated with QM methods. The electrostatic potential is then calculated based on Coulomb's law with the two charges q_i and q_j at distance r_{ij} and with ϵ_0 the dielectric permittivity of vacuum. The second term in Eq. (3.5) describes van der Waals interactions arising from dipole interactions in a molecule. These are modeled with the Lennard-Jones potential consisting of both short range repulsion and attraction terms with ϵ_{ij} being the depth of the potential well and σ_{ij} the finite distance at which the inter-particle potential vanishes. In both non-bonded energy terms r_{ij} is the distance between the particles,

$$U_{\text{non−bonded}} = \sum_i \sum_{j \neq i} \frac{q_i q_j}{4\pi \epsilon_0 r_{ij}} + \sum_i \sum_{j \neq i} 4\epsilon_{ij} \left((\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^6 \right) . \qquad (3.5)$$

A special treatment for long-ranging electrostatic interactions in MD simulations is necessary to reduce computational costs, since the electrostatic interactions decrease slowly with $1/r_{ij}$ in contrast to van der Waals interaction. Simulations are conducted using periodic boundary conditions (PBC). PBC means that the simulation box is copied an infinite amount of times by translation, and the duplicates are placed in the adjacent areas of the original box to form a continuous lattice. In this way, particles move identically in every three-dimensional periodic image (see Fig. 3.2a). During simulation, the number of particles is held constant, except for grandcanonical approaches. Therefore, once a particle leaves the simulation box at one side, one of its images from an adjacent cell will enter the box from the opposite side with identical velocity. This identical reproduction of the simulation box helps to avoid boundary artifacts, e.g., lateral particle motion at the box walls, and makes it possible to emulate in-bulk behavior without explicitly having to simulate the whole bulk. The Particle Mesh Ewald (PME) method [Darden et al., 1993] exploits the periodic structure and addresses the challenge of computationally intensive electrostatic
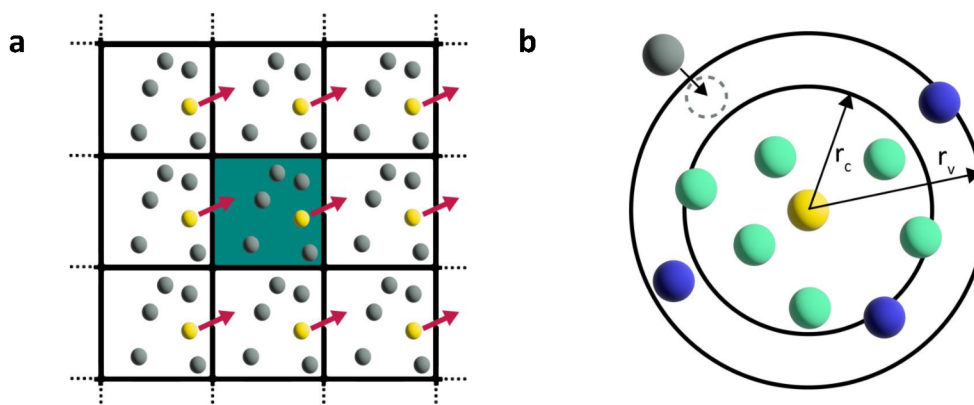
14

Figure 3.2: **Schematic representation of periodic boundary conditions and cutoff. a,**
Periodic boundary conditions: The unit cell of the system is shown in green and the mirror images
are shown in white. The particle (yellow) leaves the simulation box (green) and re-enters on the
opposite side of the simulation box with the same velocity and direction of motion (red arrow).
b, Cutoff: For the yellow particle, the interactions are only computed with all particles that are
closer than the cutoff distance r_c (green particles). All particles that are within the Verlet cutoff
r_v (both green and blue particles) are registered in the neighbor list. If particles move closer or
further away crossing the cutoff distance r_c in the course of the MD simulation, their interactions
with the yellow particle are either included or dismissed, respectively. Initially, the gray particle is
not accounted for in the neighbor list. After moving closer to the yellow particle and crossing the
r_v cutoff, it is included, once the neighbor list is updated.

interactions by separating the total electrostatic interaction into short-range and long-range
contributions. The short-range contribution is calculated directly, whereas the long-range
contribution is evaluated from a mesh of charge densities using a Fast Fourier transform
algorithm, allowing rapid convergence. Combining the two contributions yields the desired
total result in a comparably short computational time.

By contrast, Lennard-Jones interactions decrease sufficiently fast with distance that a
cut-off can be defined to save computational expense. This cutoff r_c defines the maximal
distance between two atoms at which the atoms are still included in the calculation of
van der Waals interactions. For each atom, the nearby atoms within this cutoff are added
to a neighbor list, and only for these atoms the non-bonded interactions are determined.
Since the neighbor list is usually only updated upon every tenth calculation step due to
computational capacity, sudden crossings of additional atoms from the excluded region into
the included region defined by the cutoff can occur, resulting in a jump in potential. To
prevent this, another threshold, the so-called Verlet cutoff r_v, is implemented in practice,
e.g., in the MD simulation software Gromacs, which is used in this thesis. The Verlet cutoff
is larger than the original cutoff and keeps track of all atoms approaching the non-bonded
interaction cutoff (see Fig. 3.2b).

3.1.2 Small molecule force field parameterization

Accurate force field parameters are crucial for the quality of the simulation. The parameters
are numerical values that are needed for the mathematical functions of the atom interactions
as described in Chapter 3.1.1.

15

In this thesis, the parameters of the well-optimized and validated AMBER force field [Cornell et al., 1995] are used. This force field is widely applied for simulating biomolecular systems in aqueous solution [Ponder and Case, 2003]. The AMBER force field contains parameters describing interactions of common biomolecules, like proteins. However, like in all MD force fields, parameters for heteromolecular ligands such as substrates, co-factors and potential drug molecules are not available and need to be implemented for each specific use case.

Proteins can be built from a limited set of parameters for amino acids due to their modular nature. In contrast, small molecules have vastly diverse structures and chemical properties. An estimate of the chemical space indicates 10^{60} different compounds [Bohacek et al., 1996, Dobson et al., 2004]. Even though the biologically relevant chemical space is only a fraction of the chemical space [Dobson et al., 2004], a single parameter set still cannot adequately describe such a large number of compounds. For the AMBER force field, the General AMBER Force Field (GAFF) [Wang et al., 2004] addresses this challenge with the development of a library of force field parameters for a wide range of organic molecules composed of H, C, N, O, S, P, and halogens. It contains 57 atom types with different element type, aromaticity, hybridization, and chemical environment. Atom types are transferable and can be assigned to atoms of different molecules, if the atom has the same chemical properties and a similar chemical environment. This approximation and generalization makes GAFF so efficient. GAFF is able to parameterize unique structures by averaging and extrapolating parameters from a set of *ab initio* structures.

The GAFF force field contains van der Waals and bonded parameters for each atom type. The van der Waals parameters (σ and ϵ) are taken from the AMBER parm94 [Cornell et al., 1995] or parm99 [Wang et al., 2000] force fields. The equilibrium bond length r_{eq} and the equilibrium angle θ_{eq} are derived from AMBER protein force fields, crystal structures, and *ab initio* calculations. Torsional angles are derived by fitting to quantum mechanical torsional energy profiles.

In contrast to the van der Waals and bonded force field parameters for ligands that can directly be taken from GAFF, the partial charge of each atom needs to be determined for each molecule. AMBER and GAFF force fields are not polarizable and use fixed point charges. For the charge determination, the electrostatic potentials can be derived from the quantum mechanical software suite Gaussian [Frisch et al., 2009] at the HF/6-31G* level using the restrained electrostatic potential (RESP) methodology [Bayly et al., 1993, Cornell et al., 2002]. In this way, GAFF models are compatible with other force fields of the AMBER family, since AMBER force fields use the same approach for atomic charge determination. The procedure of assigning existing force field parameters from the library to the ligand of interest and calculating its partial charges was streamlined by the development of the wrapper script ACPYPE [Da Silva and Vranken, 2012] around the ANTECHAMBER software [Wang et al., 2001] used for GAFF.

3.1.3 Integration algorithms

MD simulations of atomic trajectories over time are based on integrating the equations of motion of the atoms. The equations of motion of N particles can be written as

$$\frac{\mathrm{d}^2 \vec{r}_i}{\mathrm{d}t^2} = \frac{\vec{F}_i}{m_i}, \quad i = 1, \dots, N, \tag{3.6}$$

where the applied force $\vec{F}_i(t)$ on particle i with mass m_i at time t can be computed by the negative gradient of the potential energy U at the position of the particle,

$$\vec{F}_i = -\vec{\nabla}U(\vec{r}_i). \tag{3.7}$$

After plugging in the initial coordinates of each atom into the MD software and selecting a force field defining the potential $U(\vec{r}_i)$, the overall force on each atom at time t is evaluated as a vector sum from its interactions with other atoms (Eq. (3.7)). The required initial velocities are commonly assigned at random according to the Maxwell-Boltzmann distribution. The force on each particle determines the new positions evolved in time according to Eq. (3.6). These calculations (Eq. (3.6), Eq. (3.7)) are then alternately repeated throughout the simulation until the desired number of iteration steps is obtained.

The equations of motion in a many-particle system can only be solved numerically with discrete time steps. Various algorithms aim to integrate the equations of motion via Taylor expansion around a given time t. Because only a limited number of Taylor expansion terms are employed, this procedure introduces errors. In many MD simulation software engines, the underlying numerical integration algorithm is based on the Verlet algorithm, which has both low integration error and high computational efficiency. The Verlet algorithm is a time-symmetric integrator and uses the expansion of the particle position forward and backward in time. Two third-order Taylor expansions at position $r_i(t)$ are set up, one of them evolving in time running forward $(t + \delta t)$, while the other one evolves in time running backward $(t - \delta t)$, where $\vec{v}(t)$ represents velocity,

$$\vec{r}_i(t + \delta t) = \vec{r}_i(t) + \vec{v}_i(t)\delta t + \frac{\vec{F}_i(t)}{2m}\delta t^2 + \frac{\dddot{\vec{r}}_i}{3!}\delta t^3 + O(\delta t^4), \tag{3.8}$$

$$\vec{r}_i(t - \delta t) = \vec{r}_i(t) - \vec{v}_i(t)\delta t + \frac{\vec{F}_i(t)}{2m}\delta t^2 - \frac{\dddot{\vec{r}}_i}{3!}\delta t^3 + O(\delta t^4), \tag{3.9}$$

The sum of these two equations leads to:

$$\vec{r}_i(t + \delta t) \approx 2\vec{r}_i(t) - \vec{r}_i(t - \delta t) + \frac{\vec{F}_i(t)}{m}\delta t^2, \tag{3.10}$$

with an error of order δt^4.

While the positions of the atoms can be approximated to a certain degree of accuracy, the velocities cannot be obtained directly. The basic Verlet algorithm does not incorporate velocities explicitly. This disadvantage is overcome with the leap-frog algorithm, which is a modification of the Verlet algorithm and the default integration algorithm employed in the Gromacs package. With the leap-frog algorithm, both the positions and velocities can

be assessed directly. It computes the velocities at half-time intervals and uses positions at time t and velocities at time $(t + 1/2\ \delta t)$ to derive the updated positions $\vec{r}(t + \delta t)$,

$$\vec{r}(t + \delta t) = \vec{r}(t) + \delta t \vec{v}(t + \frac{\delta t}{2})\,, \tag{3.11}$$

$$\vec{v}(t + \frac{\delta t}{2}) = \vec{v}(t - \frac{\delta t}{2}) + \delta t \frac{\vec{F}(t)}{m}\,, \tag{3.12}$$

with an error of δt^4 in the positions.

The choice of the simulation time step is essential to ensure numerical stability. The time step should be significantly shorter than the fastest motion occurring in the system. In biological systems, the fastest movement is the bond vibration of a hydroxyl group with a stretching vibration frequency of 3600 cm^{-1}, leading to a time period of around 10 fs. Therefore, the time step should be set to 1 fs in non-constrained MD simulations. To increase computational efficiency, covalent bonds involving hydrogens are usually constrained to their ideal value. In Gromacs, O-H bonds in water models are constrained with the SETTLE algorithm [Miyamoto and Kollman, 1992], while in biomolecules, the P-LINCS algorithm is commonly deployed for covalent bonds involving hydrogens [Hess et al., 1997, Hess, 2008]. This allows us to increase the time step to 2 fs, since the fastest motion without hydrogens is the stretching vibration of a carboxyl group with a time period of 20 fs. In this thesis, a time step of 2 fs is applied.

3.1.4   MD simulation procedure

The first step before starting an MD simulation is to set up a system in a simulation box in which each atom is assigned to a position coordinate. All systems in this thesis are proteins solvated in water, and ions. Prior to initiation of an MD simulation, the system has to be energy-minimized and equilibrated. The minimization and equilibration steps are necessary to remove large steric clashes, e.g., overlapping atoms, or close contacts within and between structures. Without these pre-processing steps, energetically unfavorable structures would result in large forces and energy differences and consequently lead to termination in early MD steps. In this thesis, the steepest descent algorithm for minimization is applied, which gradually moves the atomic coordinates along the negative potential energy gradient direction until a local minimum is found. To ensure a stable simulation, an additional equilibration phase is needed. The equilibration is a short MD run with positional constraints on the heavy atoms of the protein, which can be decreased with time. The constraints keep the protein atoms near their initial positions, while the solvent molecules can move freely. The equilibration should at least last until the potential energy converges and does not change significantly anymore. If this is the case, the production run can be started.

3.1.5 Thermodynamic ensembles

From the point of view of statistical mechanics, an MD simulation can be seen as a tool to sample from a thermodynamic ensemble. The integration of Newton's equations of motion produces the microcanonical ensemble, also referred to as NVE ensemble, where the number of particles (N), the total energy (E) and the volume (V) are conserved. For biomolecular systems, a different approach called the isothermal-isobaric (NPT) ensemble is commonly used, where instead of the volume and total energy the pressure and temperature are held constant. Since NVE ensembles in practice can suffer from energy drifts over time and most experiments of biomolecules are performed at quasi-constant temperature and pressure, the NPT ensemble is the preferred choice and applied throughout this thesis. The NPT ensemble requires barostat and thermostat algorithms to respectively maintain the pressure P and temperature T at constant physiological values. The temperature and pressure values of the thermostat and barostat can be chosen and are constant by construction, while instaneous temperature and pressure of the system do fluctuate and are computed from the velocities and from interactive forces between particles, respectively.

Thermostat algorithms ensure constant temperature averages $\langle T \rangle$ by exchanging kinetic energy with the system. Several realizations of a thermostat exist. The Berendsen thermostat [Berendsen et al., 1984] for instance randomly assigns additional energy to all energy terms from the correct statistical distribution. It couples the instantaneous temperature $T(t)$ of the system to a target temperature T_0 by a relaxation process:

$$\frac{dT(t)}{dt} = \frac{T_0 - T(t)}{\tau_T} \,, \tag{3.13}$$

with the coupling constant τ_T. After the velocity and position updates at each time step, the velocities are rescaled according to Eq. (3.13). The drawback of this approach is that the total kinetic energy can shift apart between high-frequency and low-frequency energy terms, which can result in highly non-physical behaviors of the system. The Andersen thermostat [Andersen, 1980] circumvents this problem by instead changing the velocity of particles randomly. Their momenta are stochastically re-setted by a collision with the heat bath, providing a Maxwellian distribution. This method is reasonable for computing equilibration properties, but is not a reliable option, if dynamical properties are of interest. The Nosé-Hoover thermostat [Hoover, 1985, Nosé, 1984], which is applied in this thesis, introduces an additional virtual variable, which decelerates or accelerates particles until the desired temperature is obtained. This is realized by including an additional degree of freedom to the heat bath in the Hamiltonian:

$$\mathcal{H}(q, s, p, p_s) = \mathcal{H}\left(q, \frac{p}{s}\right) + \frac{p_s^2}{2Q} + gk_B T \ln(s) \,, \tag{3.14}$$

where g is the number of independent momentum degrees of freedom, k_B the Boltzmann constant, s is the variable of the additional degree of freedom, p_s its corresponding momentum, T its temperature and Q an imaginary mass.

Barostat algorithms maintain the pressure of the system by scaling the simulation box, because the volume is inversely related to pressure by Boyle's law. In MD simulations,

usually the Berendsen [Berendsen et al., 1984] or the Parrinello-Rahman [Parrinello and Rahman, 1981] barostats are employed. The Berendsen barostat is used in this thesis for equilibration runs, since it is stable even when the system pressure deviates significantly from the desired pressure. However, it does not reproduce the correct thermodynamic ensemble due to errors in the fluctuations. In contrast, the Parrinello-Rahman barostat can be unstable, if there is a large difference between system and target pressure, but does reproduce the correct thermodynamic ensemble. Thus, the Parrinello-Rahman barostat is used in this thesis for production simulations.

For the Berendsen thermostat, the pressure follows a first-order kinetic relaxation towards a given reference pressure P_0:

$$\frac{dP(t)}{dt} = \frac{P_0 - P(t)}{\tau_P} \,, \tag{3.15}$$

where τ_P is the coupling constant. For the Parrinello-Rahman barostat, similar to the Nosé-Hover thermostat, an extended Hamiltonian is used.

3.2   SWISH simulations

Rough energy landscapes of biomolecular systems with many local minima separated by high-energy barriers can limit the sampling efficiency of MD simulations. Often, not all relevant conformational states can be reached with MD simulations. Enhanced sampling methods like the below described SWISH technique or Metadynamics, described in the following section, can address this problem.

To enhance conformational sampling of cryptic pocket openings, Oleinikovas et al. [2016] developed the method called Sampling Water Interfaces through Scaled Hamiltonians (SWISH), which is based on a replica exchange MD (REMD) framework. In general, enhanced sampling methods are necessary to observe rare events such as cryptic pocket openings. Sampling transitions between different conformations that are separated by high-energy barriers with unbiased MD simulations is highly unlikely. The integration time step in MD simulations is limited to femtoseconds (Chapter 3.1.3.) and conventional MD simulations can usually not reach a millisecond, while the observation of events can exceed the timescale of seconds.

One of the most widely used and the most successful methods to enhance conformational sampling is the REMD method, which is the basis of SWISH simulations. The algorithm is based on the parallel tempering method [Geyer et al., 1991] and was first applied to MD simulations by Sugita and Okamoto [1999]. In particular, there are two types of REMD methods: Temperature (T) and Hamiltonian (H) REMD simulations. In T-REMD or H-REMD several copies of the system are simulated simultaneously and independently at different temperatures or force fields, respectively. At preset time intervals, pairs of neighboring replicas attempt to randomly exchange with a probability according to a Metropolis acceptance criterion (schematically shown in Fig. 3.3).

In T-REMD the probability for an exchange of states is:

$$P(\text{state}_1 \leftrightarrow \text{state}_2) = \min\left(1, \exp\left[\left(\frac{1}{k_B T_1} - \frac{1}{k_B T_2}\right)(U_1 - U_2)\right]\right), \qquad (3.16)$$

with T_1,T_2 and U_1,U_2 being the temperatures and the potential energies of replicas 1 and 2, respectively. Each attempted exchange of two replicas is based on the probability according to Eq. (3.16). If the exchange is accepted, the velocities are re-scaled with $(T_1/T_2)^{\pm 0.5}$.

In H-REMD the probability for an exchange of states can be rewritten to:

$$P(\text{state}_1 \leftrightarrow \text{state}_2) = \min\left(1, \exp\left[-\frac{1}{k_B T}\left((U_1(r_2) - U_1(r_1)) - (U_2(r_2) - U_2(r_1))\right)\right]\right),$$
$$(3.17)$$

with the energy difference between neighboring configurations using the force field for replica 2 minus the same difference using the force field for replica 1 [Meli and Colombo, 2013].

The random walk in temperature or Hamiltonian allows for sampling various conformational states that are trapped locally at low simulation temperatures, because states can be exchanged with replicas at higher simulation temperature or at a different Hamiltonian. These swaps between the replicas do not perturb the canonical distribution of states in each replica. With T-REMD it is possible to sample more conformations than with the same number of classical MD simulations running at the reference temperature, if the energy barriers between states are higher than the thermal energy per degree of freedom [Zuckerman and Lyman, 2006]. However, transitions with barriers that are mostly entropic are unlikely to be sampled with T-REMD [Nymeyer, 2008]. H-REMD is a common alternative and can have better convergence behavior for large systems. The sampling efficiency of T-REMD or H-REMD is determined by the aggregate number of transitions between system states (here open and closed substrate pockets) sampled across the replicas for a given investment of computational resources [Rosta and Hummer, 2009].

The number of accepted exchanges of states in REMD simulations is very important. A sufficient number of transitions of replica states is necessary to enable the lowest replica to benefit from the other replicas with higher temperatures or modified force field parameters. Each state should visit every replica as frequently as possible, which requires a sufficient overlap of sampled potential energies. The number of needed replicas is related to the difference between the ensembles.

In T-REMD simulations, the replicas vary in their physical temperature as controlled by the respective thermostats, whereas the replicas in the H-REMD simulations can be biased in an arbitrary manner to accelerate sampling. SWISH simulations are a specific type of H-REMD simulations where the developers aimed to replicate the effect of small ligands to open up hydrophobic pockets. In SWISH simulations, the replicas differ in their water-protein interaction. The non-bonded interactions of solvent molecules with the apolar protein atoms, carbon and sulfur, are increased with each replica. By this means, water molecules can mimic ligand-like behavior to increase cryptic site opening. The method makes use of a linear scaling factor λ, which can be used to increase the water affinity to
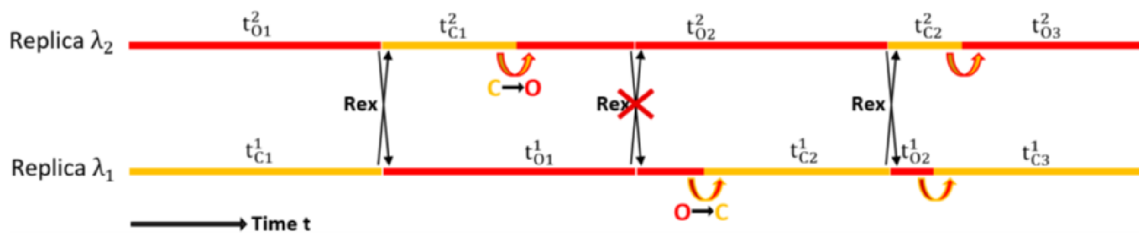
Figure 3.3: **Schematic representation of a two state SWISH simulation with protein pocket openings.** The two states of the cryptic pocket are represented as open O (red) and closed C (yellow). The parallel running replicas with $\lambda_2 > \lambda_1$ try to exchange at pre-defined time points, indicated by the black arrows (Rex). If the potential energy difference is too large, as shown with the crossed out arrow, the replicas are not exchanged. State transitions, i.e., pocket openings C $\rightarrow$ O and closings O $\rightarrow$ C, may occur within a replica indicated by the curved arrows. Time periods are labeled as $t^{\beta}_{\alpha i}$ with αi being the respective state in each replica β. Image by Tiede [2021], which was inspired by Stelzl and Hummer [2017].

the protein surface. Strong water-protein interaction leads to protein unfolding, hence the range of λ has to be chosen carefully. SWISH in its original implementation used eight replicas with a λ range of 1.0 to 1.35 [Oleinikovas et al., 2016]. In a more recent study, the developers were able to reduce the number of replicas running in parallel to six, which increased the computational efficiency [Comitani and Gervasio, 2018].

As described in Eq. (3.1) the total potential energy of a system is a sum of bonded and non-bonded interactions. In SWISH simulations, the non-bonded interaction term is scaled, while the bonded interactions remain untouched:

$$U(r_i; \lambda) = U_{\text{bonded}}(r_i) + U_{\text{non-bonded}}(r_i; \lambda). \tag{3.18}$$

Further, the components of the non-bonded interactions can be separated into water-water (ww), solute-solute (ss) and water-solute (ws) terms and only the water-solute interactions are scaled:

$$U_{\text{non-bonded}}(r_i; \lambda) = U^{\text{ww}}_{\text{non-bonded}}(r_i) + U^{\text{ss}}_{\text{non-bonded}}(r_i) + U^{\text{ws}}_{\text{non-bonded}}(r_i; \lambda). \tag{3.19}$$

Looking at the water-solute interactions, only the interactions with the apolar fraction of the solute is modified:

$$U^{\text{ws}}_{\text{non-bonded}}(r_i; \lambda) = \lambda \cdot U^{\text{WSapolar}}_{\text{non-bonded}}(r_i) + U^{\text{WSpolar}}_{\text{non-bonded}}(r_i). \tag{3.20}$$

Hence, λ only affects a small part of the non-bonded interactions. As a result, the differences in potential energy are small, which in turn is advantageous for an exchange between two replicas.
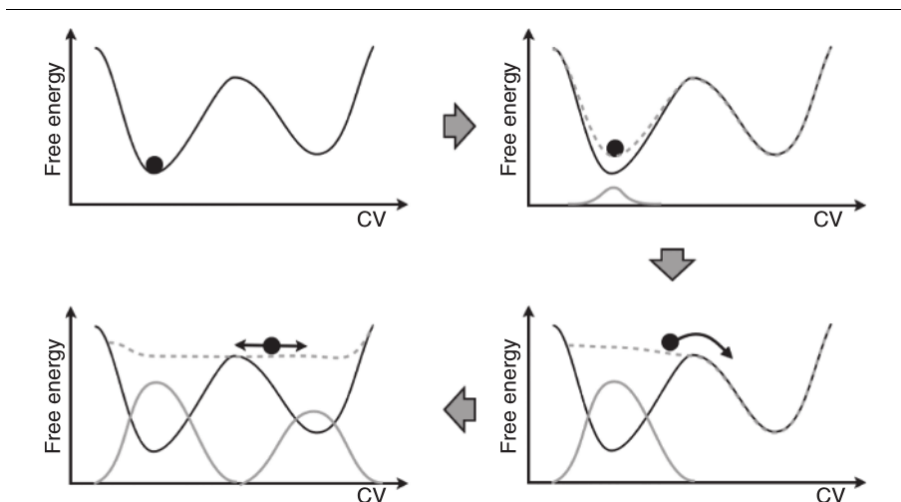
Figure 3.4: **Schematic representation of metadynamics.** In the first image, the system is simulated according to unbiased dynamics. In the second image, a Gaussian potential is added (solid gray line) that alters the free-energy landscape and raises the system (dashed gray line). In the third image, the system transitions into the other free-energy minimum state after the former minimum is "filled up" with a sum of Gaussian biasing potentials. In the fourth image, both free-energy basins are filled, and the system can evolve on a flat landscape. An estimate of the negative free-energy profile can be derived from the summation of the deposited bias (solid gray profile). Reprinted from [Bussi et al., 2015], with kind permission by John Wiley & Sons.

3.3 Metadynamics simulations

Metadynamics is an enhanced sampling technique that uses a repulsive, history-dependent biasing potential along certain collective variables (CVs). CVs are any functions of internal coordinates. The applied CVs should be good reaction coordinate models, which both allow for distinguishing between all interesting different states of the transition, also called order parameters, and for reflecting the correct kinetics. The idea of metadynamics is to fill the free energy minima with an external biasing potential also described as "filling the free energy wells with computational sand" [Laio and Parrinello, 2002]. With the help of biasing potentials, the system is pushed away from its local free energy minima, which enables faster sampling of conformational space (see Fig. 3.4). The bias is applied along pre-defined CVs. In contrast to SWISH simulations, the identification of specific CVs that are critical for sampling the rare event of interest, is necessary before starting a metadynamics simulation.

In metadynamics, the sampling is accelerated with an external history-dependent biasing potential. The biasing potential, which is a function of the CVs, is imposed on the system. It is constructed as a sum of Gaussian kernels added along the trajectory in CV space. The biasing potential $V(\vec{s}, t)$ can be expressed as:

$$V(\vec{s}, t) = \sum_{k\tau < t} W(k\tau) \exp\left(- \sum_{i=1}^{d} \frac{\left(s_i - s_i(\vec{r}(k\tau))\right)^2}{2\sigma_i^2} \right). \tag{3.21}$$

with $\vec{s}(\vec{r})$ being the CVs, τ the Gaussian deposition rate, σ_i the width of the Gaussian for the ith CV, and $W(k\tau)$ the height of the Gaussian at simulation time $k\tau$. In the long time limit, the biasing potential is the negative free energy as a function of the CVs. Hence, the

biasing potential is an estimator of the free energy $F(\vec{s})$:

$$V(\vec{s}, t \to \infty) = -F(\vec{s}) + C \, . \tag{3.22}$$

where C is a constant.

In practical use, standard metadynamics has two well-known problems. One problem is that it can push the simulated system toward non-physical high-energetic states, because the biasing potential overfills the free-energy profile. The second problem is that the estimate for the free-energy landscape does not converge but fluctuates, and thus it is not clear when to stop the simulation. To alleviate these challenges, well-tempered metadynamics was introduced in Barducci et al. [2008]. Here, the Gaussian height W decreases exponentially with simulation time according to:

$$W(k\tau) = W_0 \exp\left(- \frac{V\big(\vec{s}(\vec{r}(k\tau)), k\tau\big)}{k_B \Delta T} \right) . \tag{3.23}$$

where W_0 is the initial Gaussian height and ΔT is a parameter with the dimension of a temperature. In this way, by rescaling the Gaussian height, the biasing potential converges more smoothly in the long time limit. However, it does not completely compensate the underlying free energy:

$$V(\vec{s}, t \to \infty) = -\frac{\Delta T}{T + \Delta T} F(\vec{s}) + C \, . \tag{3.24}$$

T denotes the temperature of the system, and the input parameter ΔT can be selected to tune the extent of the free-energy exploration. In the case of $\Delta T = 0$, the simulation corresponds to standard MD, while $\Delta T \to \infty$ reproduces standard metadynamics. For well-tempered metadynamics, the input parameter is called "bias factor" γ:

$$\gamma = \frac{\Delta T}{T + \Delta T} \, . \tag{3.25}$$

γ needs to be chosen carefully, since it should be small enough for producing physical states and should be large enough for the relevant free-energy barriers to be crossed. By adjusting the bias factor, well-tempered metadynamics is able to enhance the fluctuations of the CVs in a controlled manner.

S1 pocket plasticity in trypsin-like proteases

4.1 Introduction

Trypsin-like serine proteases from family S1 constitute the largest protease group in humans [Goettig et al., 2019]. Members of this family are enzymes responsible for blood coagulation, digestion, fibrinolysis, fertilization, development, apoptosis and immunity. Because trypsin-like serine proteases serve so many important biological functions, many of them have been under active pursuit as therapeutic targets with indications ranging from thrombosis and inflammation to asthma and chronic obstructive pulmonary disease.

Trypsin-like serine proteases exhibit a highly similar tertiary folding pattern, particularly for the region near the substrate binding site comprising the conserved catalytic triad of His57, Asp102, and Ser195. The essential catalytic unit of these peptidases is a polypeptide chain of about 220 amino acid residues, which can be extended N-terminally by addition of unrelated peptide segments in many mosaic proteins [Rawlings and Barrett, 1994]. Proteolytic cleavage at the N-terminus of the catalytic domain of a proenzyme leads to a new N-terminus with a hydrophobic residue. The backbone of this residue forms a salt bridge with Asp194 resulting in the assembly of the functional catalytic site.

We studied a wide range of trypsin-like serine proteases, all of which are important drug targets: Complement factor D, cathepsin G, chymase, chymotrypsin, factor VIIa, factor IXa, factor Xa, factor XIa, factor XIIa, hepsin, matriptase, thrombin, tissue kallikrein, t-plasminogen activator, trypsin, tryptase, and u-plasminogen activator. Complement factor D, cathepsin G, chymase, and tryptase play essential roles in anti-inflammatory response. Chymotrypsin and trypsin are simple digestive proteases in the intestine, where they cleave almost any protein. Hepsin plays a role in cell growth and development. Matriptase plays a pivotol role in the formation and integrity of the intestinal epithelial barrier. Human kallikreins and kallikrein-related peptidases are expressed in nearly all tissues and fluids of the human body and participate in regulation of blood pressure. The coagulation factors VIIa, IXa, Xa, XIa, XIIa and thrombin are essential for blood clotting and in contrast to this, T-plasminogen activator and u-plasminogen activator are involved in the breakdown of blood clots.

The complex of the serine protease coagulation factor VIIa (FVIIa) with the membrane-bound tissue factor (TF) initiates the coagulation cascade leading to fibrin formation, schematically shown in Fig. 4.1. Because this coagulation cascade plays a critical role in pathological thrombosis, significant efforts have been made to design selective inhibitors of the FVIIa-TF complex as anticoagulant alternatives for the treatment of thrombotic diseases.
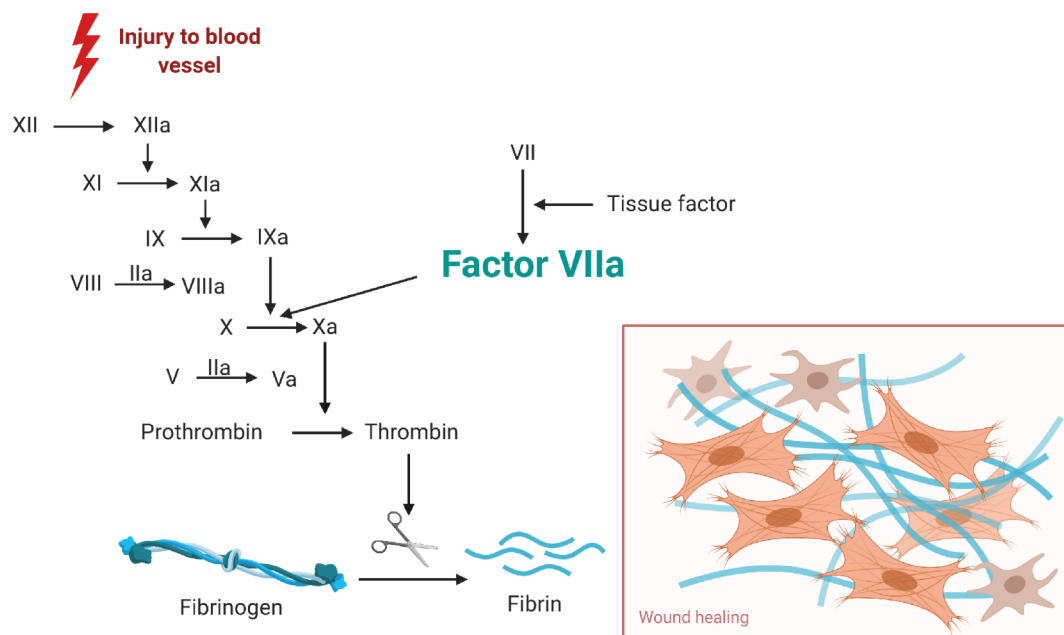
Figure 4.1: **FVIIa's role in the blood coagulation cascade.** FVIIa-TF complex activates coagulation factor X. Created with BioRender.com.

FVIIa acquires full catalytic competence only in conjunction with TF. The coagulation cascade starts with exposure of TF upon damage of the vasculature. Factor VII binds to TF and gets activated, probably autocatalytically by small amounts of enzyme in the active form FVIIa already present in circulation. In a first activation step, FVIIa is generated from the zymogen precursor FVII by specific cleavage between residues 15 and 16, resulting in a new N-terminus that ion pairs with the highly conserved Asp194 and that organizes both the oxyanion hole and primary specificity pocket. The zymogen-to-protease conversion is irreversible and renders the protease active. However, FVIIa circulating alone in the blood shows low proteolytic activity, and only becomes a fully active enzyme upon binding to the TF, which is exposed to the blood stream upon vascular injury [Banner et al., 1996, Eigenbrot et al., 2001].

Trypsin-like proteases have two major conformations: in the active conformation called E, the binding site is accessible to substrate whereas in the inactive state, E*, the binding site is blocked by the collapse of the 215-217 segment. Even after the zymogen-protease conversion has taken place, FVIIa, like other trypsin-like proteases, can adopt both active and inactive conformations. An allosteric equilibrium between active and inactive forms has been proposed early on [Dickinson et al., 1996, Eigenbrot et al., 2001]. Kinetic studies support a pre-existing equilibrium between active (E) and inactive (E*) forms for several trypsin-like proteases, including thrombin [Bah et al., 2006, Di Cera, 2008], activated protein C and factor Xa [Vogt et al., 2010]. The presence of this equilibrium in the trypsin-like proteases is also supported by thrombin structures solved both in E and E* conformation in Pineda et al. [2002, 2006]. In addition, Niu et al. [2011] demonstrated that the thrombin mutant Y225P displays both conformations in the same enzyme construct free of ligands.

All structures of FVIIa to date display the E form. The majority, 84 of 101 structures,

have a β-strand at Trp215 position and the remaining 17 structures show a flipped Trp215 side chain, but the S1 pocket is still accessible. Based on the inactive structure of the thrombin R77A mutant [Papaconstantinou et al., 2005], Pineda et al. [2006] mentioned that a collapse of Trp215 in the thrombin active site would not be possible without an initial flip of the indole side chain. Being halfway between E and E*, I call the conformation with the flipped indole E'. In all FVIIa structures published so far, the orthosteric ligands bind to the S1 pocket in a canonical fashion, all targeting the S1 pocket from the same side. Our collaborators from Sanofi reported for the first time an X-ray crystal structure of FVIIa in the E* conformation. This conformation is induced by an oxazole derivative that binds between the β-strands and that locks the protein into the inactive E* form.

With MD simulations, I was able to investigate the conformational dynamics of the S1 pocket of FVIIa at a molecular level and to sample the β-strand to loop transition at Trp215 in FVIIa, which resembles the E and E* conformations of the active site. I was able to extend these results to other serine proteases and observed similar S1 pocket flexibility and spontaneous β-strand to loop transitions in thirteen out of sixteen other serine proteases studied. This suggests that the E-E* transition with a collapse of Trp215 into the active site is a property of many serine proteases and not just thrombin or FVIIa.

4.2 Methods

4.2.1 MD simulations

The coordinates of FVIIa with the soluble TF (two Sanofi in-house structures, PDB entry: 2aer [Bajaj et al., 2006]) served as starting points for different simulation set-ups: The ligand structures of the internal structures were either kept or removed. The ligand structure of 2aer was removed. Zn^{2+} and Ca^{2+} ion coordinates were added to the internal structures from PDB entry 2aer after protein alignment using VMD [Humphrey et al., 1996]. All residues were simulated in their physiological protonation state, except His57 (chain B) and Glu130 (chain C) were protonated. Ligand structures and γ-carboxyglutamic acid side chains were parameterized with the General Amber Force Field (GAFF) [Wang et al., 2004].

The coordinates of all other serine endopeptidases were taken from the PDB (Table 4.1) and simulated in their physiological protonation state, except for doubly protonated His57 in complement factor D and for protonated Glu226 in cathepsin G and Glu70 in Factor IXa, respectively. All bound ligands were removed.

Missing side chains in all set-ups were modeled using Modeller [Šali and Blundell, 1993]. Crystallographic water molecules and ions were retained. Each set-up was simulated with two different water models. The protein structures were solvated once in TIP4P-D [Piana et al., 2015] and once in TIP3P water [Jorgensen et al., 1983] with 150 mM NaCl each. MD simulations were carried out using Gromacs 2018 [Abraham et al., 2015] and the AMBER99SB*-ILDN-q force field [Hornak et al., 2006a, Best and Hummer, 2009, Best et al., 2012, Lindorff-Larsen et al., 2010]. For FVIIa-TF systems, the Gromacs 2018 version with additional RTC implementation [Wassenaar, 2018] was used. Each system was energy minimized, followed by five equilibration steps, in which I gradually weakened

Table 4.1: Protein and PDB entries used as starting structures for MD simulations.

protein	PDB entry
cathepsin G	1cgh [Hof et al., 1996]
chymase	3s0n [Lo et al., 2011]
chymotrypsin	1ggd [Neidhart et al., 2001]
complement factor D	5nb7 [Lorthiois et al., 2017]
factor IXa	6mv4 [Vadivel et al., 2019]
factor Xa	1g2l [Nar et al., 2001]
factor XIa	3bg8 [Buchanan et al., 2008]
factor XIIa	6b77 [Dementiev et al., 2018]
hepsin	5ce1
matriptase	4jyt [Goswami et al., 2013]
thrombin	5a2m
tissue kallikrein 2	4nfe [Skala et al., 2014]
t-plasminogen activator	1rtf [Lamba et al., 1996]
trypsin	1h4w [Katona et al., 2002]
tryptase	4a6l, chain A [Liang et al., 2012]
u-plasminogen activator	1c5m [Katz et al., 2000]

the position restraints on heavy atoms, first in an NVT ensemble (0.25 ns) and then in an NPT ensemble (4 x 0.5 ns) using a Berendsen thermostat and barostat [Berendsen et al., 1984]. Production simulations were run at a temperature of 310 K and a pressure of 1 bar in an NPT ensemble using a Nosé-Hoover thermostat [Evans and Holian, 1985, Nosé, 1984] and a Parrinello-Rahman barostat [Parrinello and Rahman, 1981].

4.2.2 Metadynamics simulations

Equilibrated FVIIa structure (PDB entry: 2aer with removed ligand) solvated in TIP3P water served as starting structure for well-tempered metadynamics simulations using Gromacs 2018 and PLUMED 2 [Tribello et al., 2014]. The collective variable O_{Trp215}-N_{Val227} distance was used to simulate the S1 pocket conformational change. Gaussians were deposited every 2 ps with a starting height of 0.1 $\frac{kJ}{mol}$, which gradually decreased with a bias factor of 2 and a temperature of 310 K. The width of the Gaussians was defined on the basis of unbiased MD simulation runs and was set to 0.02 nm. Upper walls with a force constant of 2000 $\frac{kJ}{mol}$ were set to 1.2 nm, 0.4 nm and 0.4 nm for the O_{Trp215}-N_{Val227} distance, O_{Ile212}-N_{Thr229} distance and N_{Ile212}-O_{Thr229} distance, respectively.

4.3 Results and discussion

4.3.1 Ligand-bound X-ray crystal structures with oxazole and benzamidine derivative show different S1 pocket conformation

Our collaborators from Sanofi determined in total two new high resolution FVIIa crystal structures: (i) The structure of FVIIa with the benzamidine-based inhibitor. (ii) The structure of FVIIa with a bound oxazole derivative in the E* conformation with a collapsed Trp215. Both structures are of human wild-type FVIIa in complex with TF. The

structure of the benzamidine-based derivative was determined at 1.57 Å and the structure of the oxazole inhibitor at 1.87 Å. Both ligands inhibit FVIIa enzymatic activity. For the benzamidine and oxazole derivatives, K_i's of 0.017 µM and 1.7 µM were measured, respectively.

The two X-ray crystal structures are largely similar, except in the conformation of residues 215 to 220, which define one side of the S1 pocket (Fig. 4.2a,b,c), and in the loop 170D to 170G. The benzamidine-derivative-bound structure displays a β-sheet at Trp215 (E conformation), whereas the 215-217 segment is collapsed into the active site in the oxazole-derivative-bound structure (E* conformation).
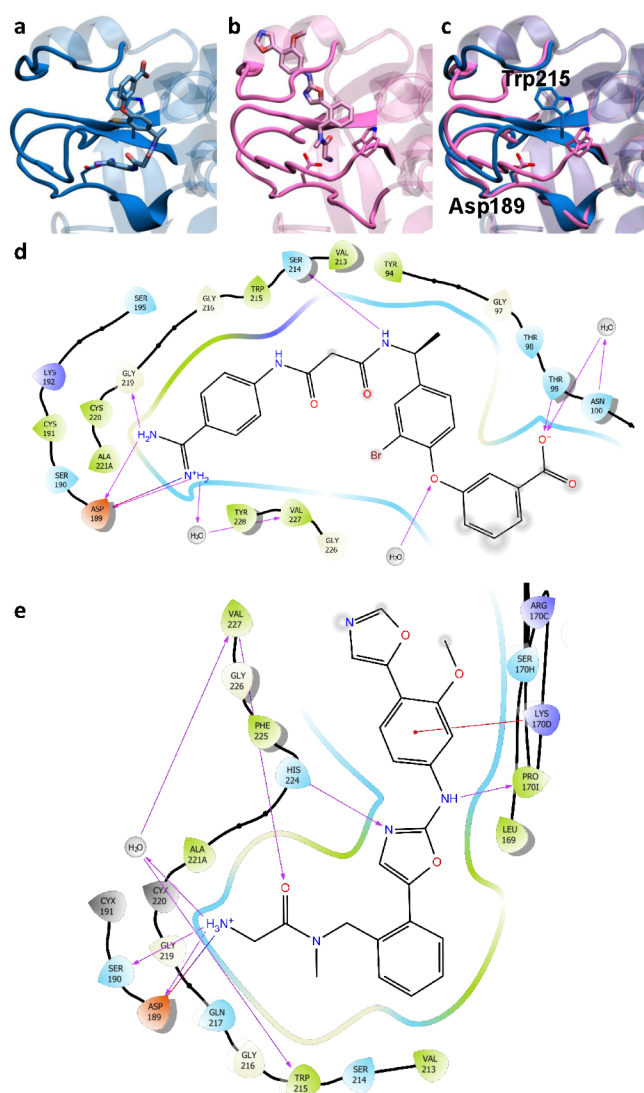

Figure 4.2: **S1 pocket in FVIIa. a,** benzamidine-derivative-bound X-ray crystal structure. **b,** oxazole-derivative-bound X-ray crystal structure. **c,** superposition of benzamidine- (blue) and oxazole-derivative-bound (magenta) X-ray crystal structures. **d,** ligand interaction diagram of benzamidine derivative. **e,** ligand interaction diagram of oxazole derivative.

The main interaction of the benzamidine derivative to FVIIa are the electrostatic interactions between the amidine group and Asp189 (Fig. 4.2d). In addition, one of the malonamide NH's is hydrogen bonded to the backbone of Ser214 while the other interacts with Ser195-Oγ. The meta-bromine from the central phenyl ring was observed to increase affinity [Klingler et al., 2004] and is positioned next to the side chain of Trp215. A para-phenoxy-group with a meta-carboxylate is pointing toward the S4 pocket. The carboxylate group has hydrogen bonds with the side chain of Thr99 and with the backbone of Asn100 via a water molecule.

In contrast to the benzamine-derivative-bound structure, the oxazole-derivative-bound structure reveals a completely new binding mode. In the benzamidine-derivative-bound structure and in all previous reported complex structures of FVIIa, the ligands target the S1 pocket from the same side. However, the oxazole derivative binds to the S1 pocket from a different direction. It wedges in between the β-strands at Trp215 and Val227 (Fig. 4.3c) and does not bind in the binding pocket
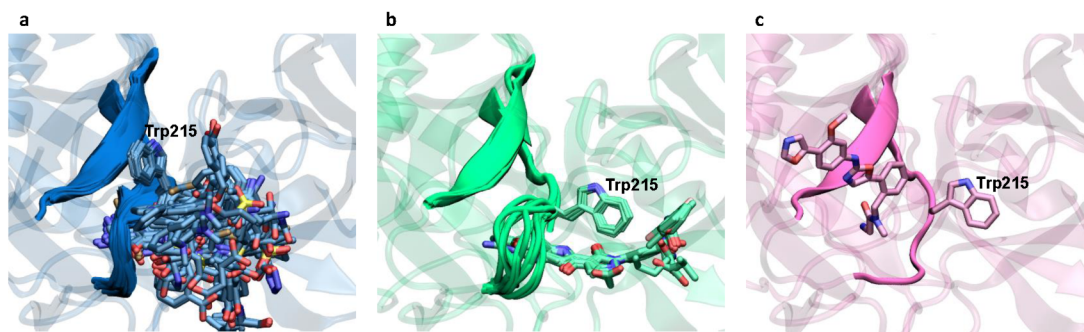
Figure 4.3: **Superimposed X-ray crystal structures of FVIIa. a,** E conformation with β-strand at Trp215: Sanofi in-house benzamidine-derivative structure and structures from PDB (Table 4.2). **b,** E' conformation with displaced Trp215 (PDB entries: Table 4.2). **c,** E* conformation: Sanofi in-house oxazole-derivative structure.

Table 4.2: Structures used to generate Fig. 4.3

PDB entries of E conformation	PDB entries of E' conformation
1cvw, 1dan, 1dva, 1j9c, 1kli, 1o5d, 1qfk, 1w0y,	2f9b, 2flr, 5pab, 5pac,
1w2k, 1w7x, 1w8b, 1wqv, 1wss, 1wtg, 1wun, 1wv7,	5pae, 5pan, 5pag, 5pai
1ygc, 1z6j, 2a2q, 2aei, 2aer, 2b7d, 2b8o, 2bz6, 2c4f,	5pao, 5pav, 5pax, 5pay,
2ec9, 2fir, 2flb, 2puq, 2zp0, 2zwl, 2zzu, 3ela, 3th2,	5pb2, 5pb3, 5pb4, 5pb5, 5pb6
3th3, 3th4, 4ibl, 4ish, 4isi, 4jyu, 4jyv, 4jzd, 4jze,	
4jzf, 4na9, 4ng9, 4nga, 4x8s, 4x8t, 4x8u, 4x8v,	
4ylq, 4yt6, 4yt7, 4z6a, 4zma, 4zxx, 4zxy, 5i46, 5l2y,	
5l2z, 5l30, 5pa8, 5pa9, 5paa, 5paf, 5paj, 5pak,	
5pam, 5paq, 5par, 5pas, 5pat, 5pau, 5paw, 5pb0,	
5pb1, 5tqe, 5tqf, 5tqg, 5u6j, 6r2w	

of natural substrates or previous reported ligands (Fig. 4.3a,b). The oxazole inhibitor stabilizes the E* conformation around the S1 pocket. It forms a similar salt bridge with Asp189 and Ser190 at the bottom of the S1 pocket as in other benzamidine-based inhibitors, but targets the acidic aspartate residue from a different direction. Apart from Asp189 and Ser190 the residues involved in protein-ligand interaction are completely different from previous structures (Fig. 4.2e). The positively charged amine group is not only interacting with Asp189 and Ser190, but also has water-mediated interactions with the protein backbone of residues Trp215 and Val227. Additionally, the amino group of Val227 forms a hydrogen bond with the carbonyl group of the ligand. Two more hydrogen bonds are present between the ligand and protein backbone of His224 and Pro170 and one π-cation interaction is formed between Lys170 and one benzene moiety of the ligand.

4.3.2 MD simulations support a pre-existing equilibrium between E and E* forms in FVIIa

To understand the conformational dynamics of the S1 pocket, I performed MD simulations of the FVIIa-TF complex. I used three different starting structures: the two ligand-bound structures discussed above and the benzamidine-bound structure from PDB entry: 2aer,
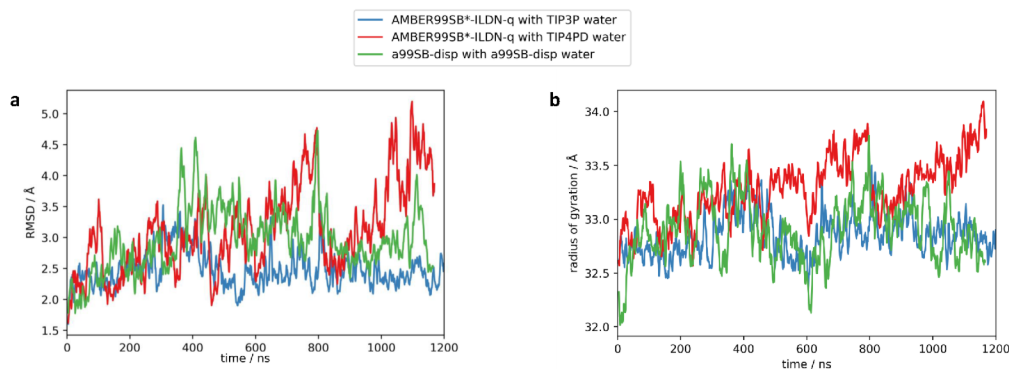
Figure 4.4: **Water force field screening of FVIIa. a,** backbone RMSD as function of time. **b,** backbone radius of gyration as function of time.

which is highly similar to the current structure with the benzamidine derivative (backbone RMSD of serine protease domain: 0.58 Å).

For both the oxazole-derivative-bound and the current benzamidine-derivative-bound structure, I conducted four MD simulations, each 4 µs long. To assess S1 pocket flexibility and complex stability, I ran the simulations in presence or absence of the ligand for each structure. For the structure from PDB entry:2aer, ten MD simulation replicates without the benzamidine ligand were performed for 500 ns each. A brief initial water force field screening of 1.2 µs showed that TIP4P-D water led to the most and TIP3P water to the least protein flexibility of FVIIa (Fig. 4.4). To bracket the effects of solvent-induced protein flexibility, I ran MD simulations of all systems both with TIP3P and TIP4P-D water.

In addition to visual inspection, the transitions between E and E* conformations were investigated by monitoring the backbone distances between Trp215 and Val227 (Fig. 4.5). Below 4 Å, backbone hydrogen bonds result in β-sheet structures. Above 4 Å, the 215-217 segment displays a loop conformation. In general, most loop-like structures at Trp215 display E* conformations (non-accessible active site), whereas all β-strand structures display E conformations (accessible active site) of trypsin-like proteases. To assess β-strand to loop transitions, both backbone distances are plotted for the 4 µs trajectories and the distance between the backbone nitrogen of Val227 and the backbone oxygen of Trp215 is plotted for the replicate simulations.

In the ligand-bound structures, the proteins remained in their respective starting conformation for both water models and inhibitors over the course of the trajectories (Fig. 4.5b), while the ligands showed different levels of flexibility depending on the water model and on the inhibitor. In the TIP3P set-up, the oxazole derivative lost the key interaction N-Asp189 quickly, rebound and left this binding site again at ≈600 ns, while it remained attached in the loop 170D to 170G. The benzamidine derivative remained bound, suggesting a more stable binding pose for the benzamidine derivative (Fig. 4.6a,c). In the TIP4P-D set-up, the bound benzamidine and oxazole derivatives both diffused slightly out of their respective binding pockets at the beginning of the MD simulations, but rebound after ≈200 ns and remained bound to Asp189 (Fig. 4.6b). However, both ligand structures in the TIP4P-D set-ups were highly flexible (heavy-atom RMSD up to 13 Å) probably due to the
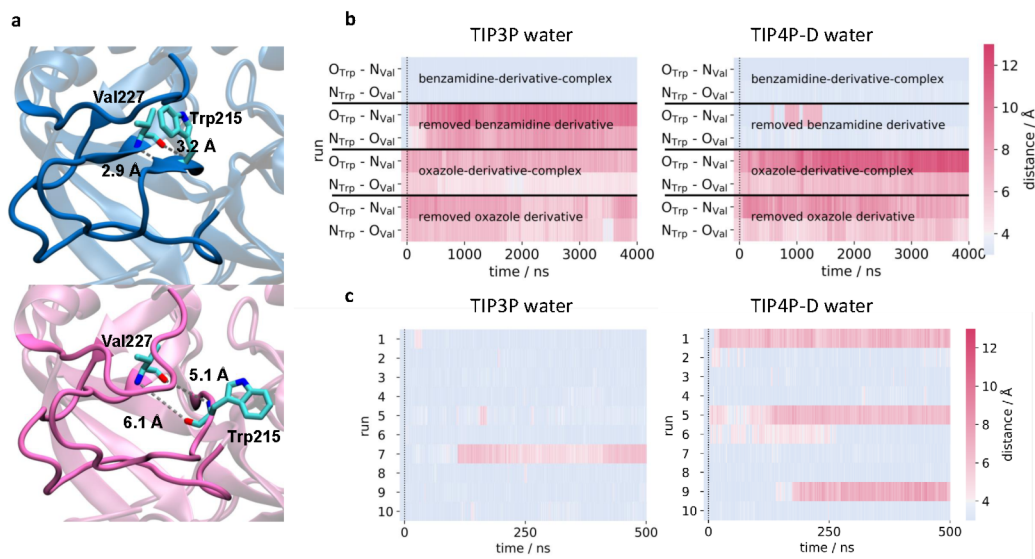
31

Figure 4.5: **Backbone hydrogen bonding in FVIIa. a,** X-ray crystal structures of benzamidine-derivative-bound structure (blue) and oxazole-derivative-bound structure (magenta) with Val227 and Trp215 shown with atomic detail. **b,** N_{Trp215}-O_{Val227} and N_{Val227}-O_{Trp215} distance as function of time for different MD set-ups as depicted. **c,** N_{Val227}-O_{Trp215} distance as function of time for 10 replicas of MD simulations with TIP3P water (left) and TIP4P-D water (right), respectively, initiated from the PDB entry: 2aer (β-strand at Trp215) structure.

destabilizing nature of the TIP4P-D water model (Fig. 4.6d).

In the simulations without bound ligands, I observed the opposite behaviour. The loop conformations persisted for both water models also without the oxazole derivative. In the 4 μs trajectories starting from the β-strand conformation with TIP4P-D water, I observed some flexibility in the strand, whereas with TIP3P water the strand transitioned quickly to the loop conformation. To acquire more data, I analyzed the 10 times 500 ns simulations with the β-strand conformation as starting structure for both water models. I observed two out of ten transitions from β-strand to loop conformation for the TIP3P water model and four out of ten transitions for the TIP4P-D water model. In run 6 with the TIP4P-D water model, I also observed re-transitioning to the β-strand conformation (Fig. 4.5c). In all transitions from β-strand to loop, water molecules probe the backbone hydrogen bonds and eventually enter between the two former β-strands. Cumulatively, the MD data suggest that the S1 pocket can undergo the conformational change between E and E* form even in the absence of ligands, in agreement with previous experimental [Bah et al., 2006, Di Cera, 2008, Vogt et al., 2010, Pineda et al., 2002, 2006, Niu et al., 2011] and computational studies [Sorensen et al., 2016, Plattner and Noé, 2015, Madsen and Olsen, 2021] on trypsin-like proteases.

In all MD simulations of FVIIa I observed high flexibility of the 170 loop (Fig. 4.7). The movement of the 170 loop is independent of the β-strand to loop transitions at Trp215 and of the presence of ligands. This is qualitatively consistent with the elevated B-factors of the 170 loop in the X-ray crystal structures. Further evidence for flexibility of the 170 loop comes from the fact the loop is not resolved in more than 40% of over 100 FVIIa
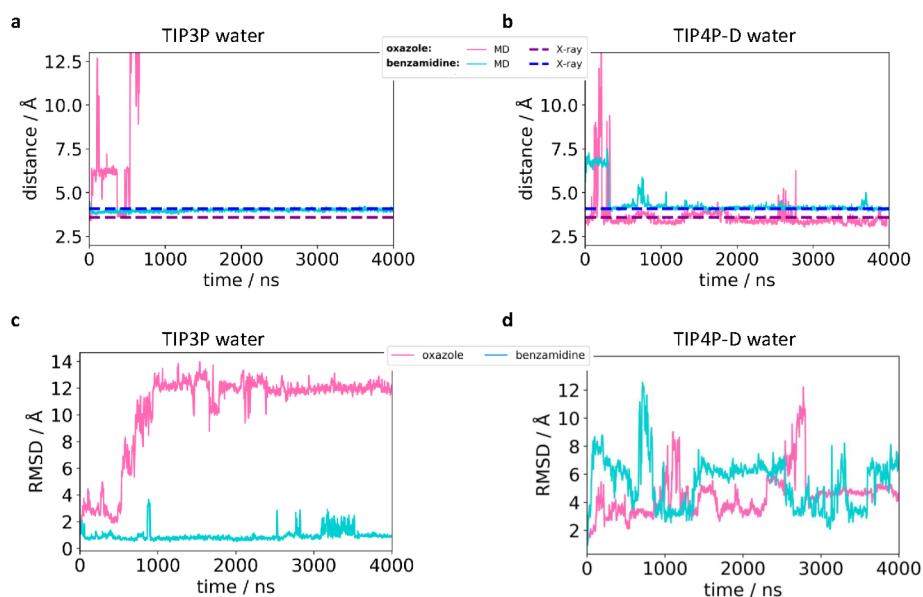
Figure 4.6: **Protein-ligand distance and ligand RMSD Interaction between Asp189 and the amino group of the ligands. a,** N_{ligand}-$C\gamma_{Asp189}$ distance as function of time in the TIP3P set-up. **b,** N_{ligand}-$C\gamma_{Asp189}$ as function of time in the TIP4P-D set-up.
Distances for benzamidine and oxazole derivatives are shown in cyan and pink, respectively. Dashed blue and magenta lines denote corresponding distances in the respective X-ray crystal structures. **Heavy-atom RMSD of the ligands from their respective minimized starting structure as function of time after superimposing the serine protease domain. c,** ligand RMSD as function of time in the TIP3P set-up. **d,** ligand RMSD as function of time in the TIP4P-D set-up. RMSD for benzamidine and oxazole derivatives are shown in cyan and pink, respectively.
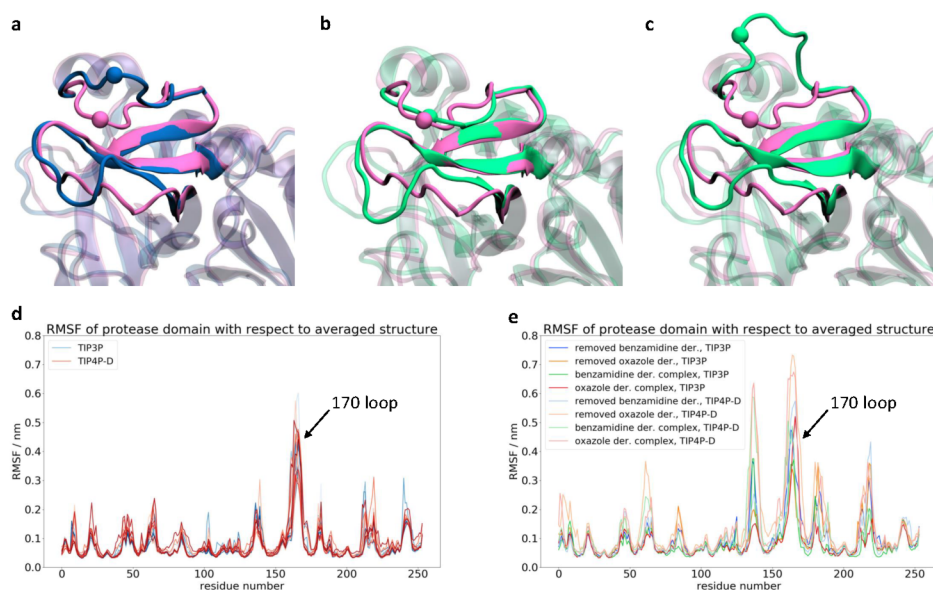
Figure 4.7: **Flexibility of 170 loop in MD simulations**. **a,** Superposition of crystal structure of FVIIa oxazole inhibitor complex (magenta) with the benzamidine structure (blue; PDB entry:2aer used as starting structure for MD) shown in cartoon. Spheres denote Gly170F positions. **b,** Superposition of crystal structure of FVIIa oxazole inhibitor complex (magenta) with MD simulation frame at 455 ns (green; run 3 of TIP3P set-up) shown in cartoon. Spheres denote Gly170F positions. **c,** Superposition of crystal structure of FVIIa oxazole inhibitor complex (magenta) with MD simulation frame at 245 ns (green; run 3 of TIP3P set-up) shown in cartoon. Spheres denote Gly170F positions. **d,** C_α RMSF of protease domain with respect to the averaged structure for 10 x MD simulations of PDB entry:2aer with removed benzamidine for both TIP3P and TIP4P-D water model. **e,** C_α RMSF of protease domain with respect to the averaged structure for different MD set-ups as depicted (der. denotes derivative).

crystal structures deposited in the PDB.

4.3.3 S1 pocket plasticity also observed in MD simulations of thirteen other serine peptidases

To assess the S1 pocket plasticity in related protein drug targets, I selected the following set of serine peptidases for MD simulations: Complement factor D, cathepsin G, chymase, chymotrypsin, factor IXa, factor Xa, factor XIa, factor XIIa, hepsin, matriptase, thrombin, tissue kallikrein, t-plasminogen activator, trypsin, tryptase and u-plasminogen activator. I used the E form ligand-free structures for all systems as the starting points, except for complement factor D. Since complement factor D is a poorly active protease stabilized in the E* form and only the binding of cofactors enhances its activity [Forneris et al., 2010], I used the E* form as starting structure. Structural alignment of all selected proteases shows the difference in the 215-217 segment. At the 215 position, complement factor D does not display a large hydrophobic side chain, unlike the other 15 proteases, and its Asp189-Arg218 salt bridge stabilizes the inactive form (Fig. 4.8). For all proteins, I set up the simulation once in TIP3P and once in TIP4P-D water.

We observed S1 pocket plasticity for thirteen of the sixteen serine peptidases studied in addition to FVIIa. Consistent with the simulations of FVIIa, I observed more transitions
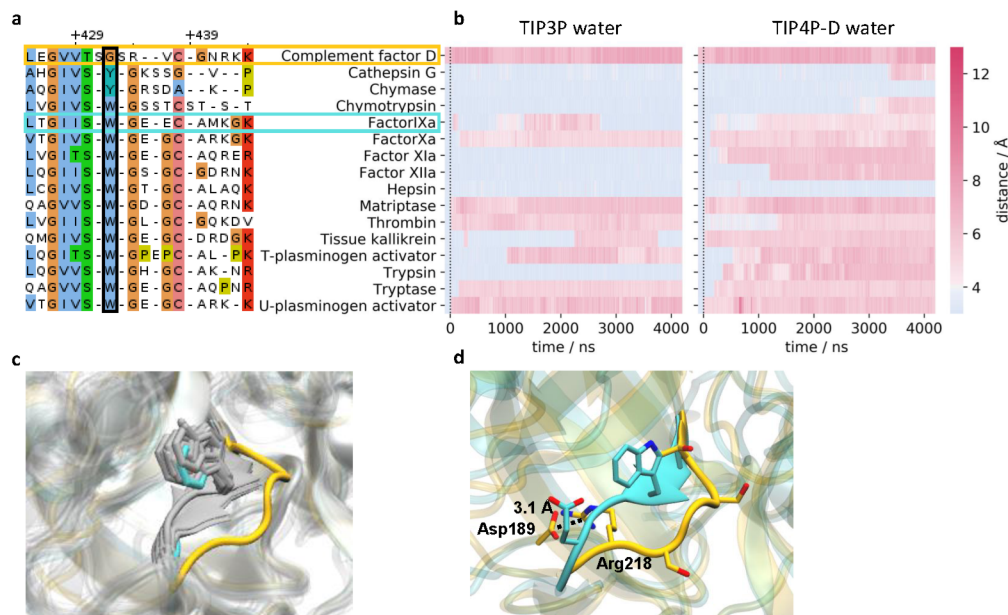
Figure 4.8: **Backbone hydrogen bonding in sixteen serine proteases. a,** partial sequences from structural alignment, black box denotes amino acid according to Trp215 position in FVIIa. **b,** backbone hydrogen bonding distance equivalent to N_{Val227}-O_{Trp215} in FVIIa as function of time from MD simulations. **c,** structural alignment of serine peptidases: complement factor D (yellow), factor IXa (cyan), all other peptidases (grey). **d,** structural alignment of complement factor D (yellow) and factor IXa (cyan).

between β-strand and loop conformations for systems solvated in TIP4P-D water than in TIP3P. This meets our expectations, since TIP4P-D can destabilize folded protein structures to some extent [Robustelli et al., 2018]. Complement factor D remained in its starting conformation for the whole trajectory of 4 μs as expected since no cofactor was added to the simulation which could displace the self-inhibitory salt bridge. Interestingly, I observed β-strand to loop transitions for all remaining proteases in our subset, except for chymase and hepsin. In the TIP3P water set-up, I also observed re-transitions to the β-strand form for factor IXa and tissue kallikrein suggesting the force field allows for sampling the equilibrium between E and E* forms.

4.3.4 Well-tempered metadynamics simulation indicated a small free energy barrier of the conformational change

The energy barrier has to be sufficiently small to observe β-strand to loop transitions with unbiased MD simulation within a microsecond time frame. However, most transitions were observed in only one direction: from β-strand to loop form. To gain insight into the energy landscape underlying the conformational change, I set up several well-tempered metadynamics simulations, in which I biased the N_{Trp215}-O_{Val227} distance. A starting Gaussian height of 0.1 kJ/mol and a bias factor of 2 led to a converged calculation. I obtained two forward, one backward transition, and ended up in an intermediate state with the O_{Trp215}-N_{Val227} distance swinging between 3 and 6 Å (Fig. 4.9a,d). In the intermediate state the hydrogen bond at N_{Trp215}-O_{Val227} was formed and the biased distance at O_{Trp215}-N_{Val227}

oscillated. Interestingly, the second transition took place without deposition of an additional Gaussian (Fig. 4.9e). This indicates that the simulation from this part onwards was run on a biased potential accumulated through the initial deposited Gaussians. Noteworthy, biasing the O_{Trp215}-N_{Val227} distance did not lead to a converged simulation. The metadynamics simulation yielded a free energy barrier of ≈ 8 kJ/mol between the favored β-strand form and the unstable loop form (Fig. 4.9b,c). The small energy barrier for the conformational change is consistent with the general concept of the existence of an equilibrium between E and E* forms. It is also interesting to consider the kinetics of the E to E* interconversion. From a Markov state model built on extensive MD simulations of trypsin [Plattner and Noé, 2015], transitions between different apo forms reminiscent of E and E* were inferred to occur on the ten-microsecond timescale, which is slower than our observations. For the related Factor Xa, stopped-flow experiments using sodium binding as a perturbant, and intrinsic fluorescence as probe showed even slower kinetics in the millisecond regime [Vogt et al., 2010]. Other than differences between the proteins, possible reasons for the more rapid sampling of E*-like states here are that I investigate β-sheet to loop interconversions, and that the E* state may not be fully formed in our MD simulations, and that the E/E* structures are differently defined in the stopped-flow experiments. Differences in MD force fields may also play a role.

4.4   Summary and conclusion

Here, using MD simulations I report spontaneous transitions between the β-strand and loop forms of the 215-217 segment in FVIIa, demonstrating S1 pocket plasticity even in the absence of ligands and with bound tissue factor. The oxazole derivative bound to FVIIa shifts the E-E* equilibrium towards the inactive conformation and stabilizes this inactive conformation. Atomistic MD simulations provided detailed insights into the pre-existing equilibrium between β-strand and loop forms at Trp215, which correlates with closed (E*) and open (E) conformations of the active site. This conformational change controls the onset of substrate binding and catalysis. In the E* form, substrates cannot bind to the active site, and catalysis is impeded. In our MD simulations, I observed multiple spontaneous transition events. In 7 out of 22 runs with the E form as starting structure, the 215-217 segment eventually rearranged into a loop structure. These transitions occurred spontaneously and were associated with water molecules probing the backbone hydrogen bonds. I observed a similar destabilizing effect of water molecules wedging into the binding interface also in the complex between the papain-like protease of SARS-CoV-2 and its di-ubiquitin substrate (see Chapter 7).

S1 pocket plasticity appears to be common among serine proteases. I observed spontaneous β-strand to loop transitions in MD simulations of 13 other serine proteases (cathepsin G, chymotrypsin, factor IXa, factor Xa, factor XIa, factor XIIa, matriptase, thrombin, tissue kallikrein, t-plasminogen activator, trypsin, tryptase, and u-plasminogen activator). Only for hepsin and chymase no transitions were observed. The results reported in this Chapter offer a view of trypsin-like proteases in the free form that is entirely consistent with the E*–E equilibrium concept. The observed plasticity of the S1 pocket in simulations
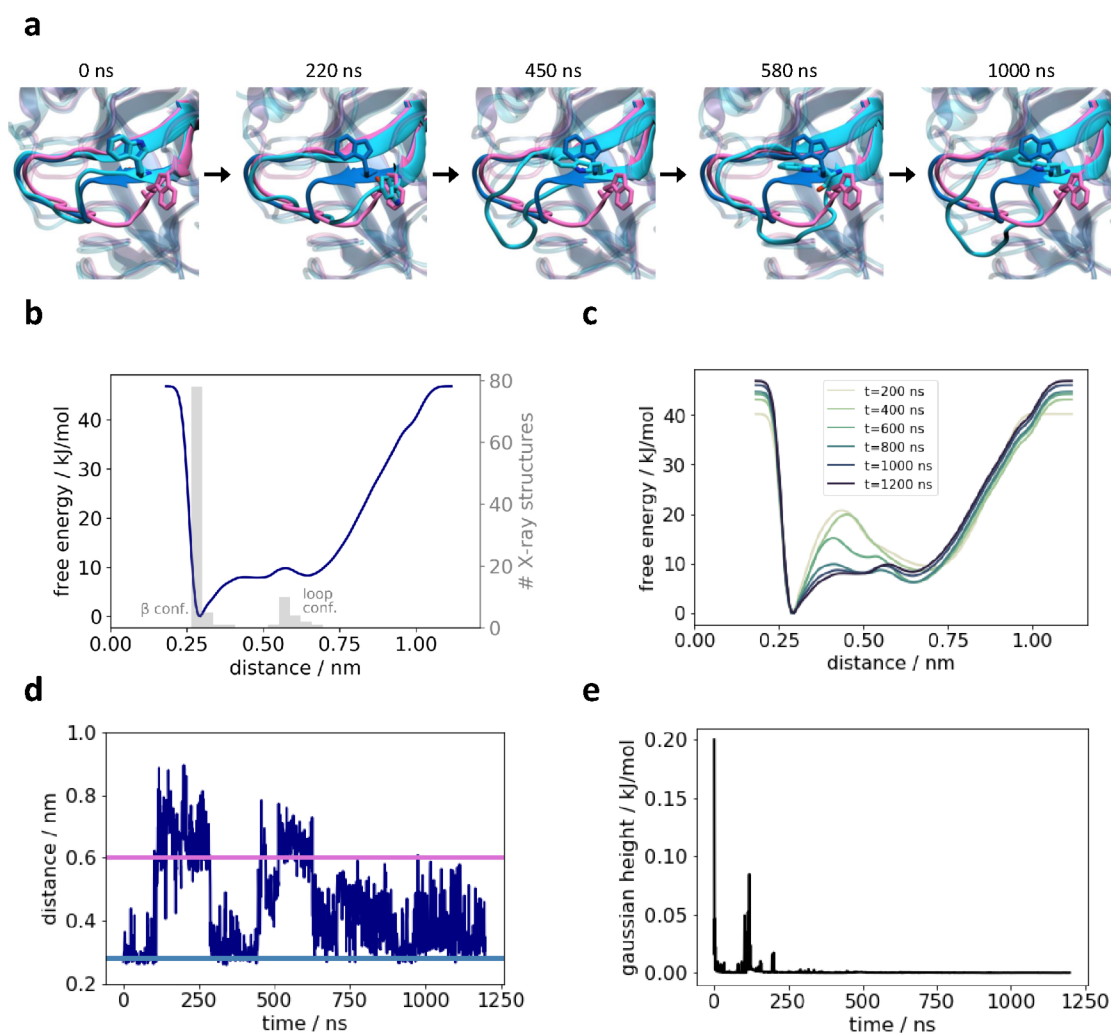
Figure 4.9: **Metadynamics simulation of FVIIa. a,** Snapshots of superpositions of metadynamics simulation frames in cyan with X-ray crystal structures of benzamidine-derivative-bound (blue) and oxazole-bound structures (magenta). **b,** free energy profile of N_{Trp215}-O_{Val227} distance with histogram of values from X-ray structures. **c,** convergence of the free energy profile of N_{Trp215}-O_{Val227} distance. **d,** O_{Trp215}-N_{Val227} distance as function of time, blue and magenta lines denote corresponding X-ray crystal structure values of benzamine-bound and oxazole-derivative-bound structures, respectively. **e,** Gaussian height as a function of time.

and the structure of the oxazole derivative bound to FVIIa in an E* conformation together suggest that the E* conformation is a valid target for the design of serine protease inhibitors. MD simulations indicate that S1 pocket plasticity can be extended to a wide range of other trypsin-like proteases.

Flap dynamics and Trp39 side chain flipping in renin

5.1 Introduction

Aspartic proteases are highly conserved from mammals to retrovirus [Revuelta et al., 2014]. Eukaryotic aspartic proteases include renin and pepsin. Renin is critical for controlling blood pressure homeostasis, and electrolyte balance as part of the renin-angiotensin-aldosterone system (RAAS). Pepsin is a digestive enzyme that degrades proteins in the stomach. Aspartic proteases play also a vital role in the lifecycle of many pathogens, including HIV-1 protease, and plasmepsin, which is found in the Malaria-causing parasite *Plasmodium falciparum.*

RAAS is an important system in the body to regulate blood pressure and the amount and composition of extracellular fluid. It consists of a sequential cascade of enzymatic reactions, schematically shown in Fig. 5.1. Specialized cells inside the kidneys from the juxtaglomerular apparatus measure blood pressure. When these cells sense a decrease in blood pressure, they produce renin, which is released into the blood. In the blood, renin cleaves angiotensinogen, produced in the liver, to generate angiotensin 1. The latter is then cleaved by angiotensin-converting enzyme (ACE) from the lungs to yield angiotensin 2. Angiotensin 2 is the strongest vasoconstriction substance in the body and activates vasoconstriction by binding to the angiotensin 1 receptor. Additionally, angiotensin 2 releases aldosterone from the adrenal cortex. Aldosterone is a hormone that increases blood pressure by reabsorption of sodium and water in the vascular system. Hence, for the prevention and treatment of hypertension and other related disorders, the most interesting inhibitory targets are the angiotensin 1 receptor, the enzymes ACE, and renin. Renin is a favorable target, since renin catalyzes the rate limiting step in RAAS and inhibition of renin prevents the formation of additional harmful angiotensin I peptides. Further, the binding site of renin has only one known substrate, implying high specificity and an opportunity for designing a small molecule causing little side effects [Gradman and Kad, 2008].

Pepsin-like aspartic proteases possess a characteristic β-hairpin structure known as the flap, which covers the active site. The eukaryotic aspartic proteases consist of two domains of a single polypeptide chain, while the active enzyme of retroviral aspartic proteases is a dimer of two identical subunits. The active site has a large cavity and is composed of two aspartic acid residues, Asp32 and Asp215, one in N- and one in C-terminal domain [Blundell et al., 1983]. Here, and throughout the thesis, pepsin numbering for residues is used. One of the catalytic Asp residues acts as proton acceptor from the nucleophilic water molecule and is assumed to be charged, while the other one is assumed to be protonated,

acting as a general acid to transfer its proton to the substrate [Andreeva and Rumsh, 2001]. The flap covering the active site can be in an open or closed conformation. Its conformational flexibility plays a crucial role in catalytic activity and substrate intake. Most aspartic proteases share structural and sequence similarity, with Trp39 and Tyr75 being conserved amino acids next to the binding site. The majority, 78 out of the 86 available X-ray crystal structures of protein-ligand complexes of renin in the PDB, show a closed flap conformation, which tightly caps the ligand molecule. Interestingly, 8 X-ray crystal structures of inhibitor-bound complexes show the flap in an open conformation, with the side chains of both Trp39 and Tyr75 flipped compared to the closed flap conformation. In addition, there are only three X-ray crystal structures available of unbound renin (PDB entries 2ren [Sielecki et al., 1989], 1bbs [Dhanaraj et al., 1992], and 2x0b [Zhou et al., 2010]), which show the flap in a closed conformation with no free space to allow substrate intake and product exit.

The flap dynamics of aspartic proteases, including renin, HIV-1 protease, BACE1 and plasmepsin II, was investigated by several experimental and computational studies [Hong and Tang, 2004, Pietrucci et al., 2009, Tzoupis et al., 2012, Roche et al., 2015]. Further, MD simulations showed that the unliganded protease can reversibly transition between closed and open flap forms [Hornak et al., 2006b, Sadiq and De Fabritiis, 2010, Xu et al., 2012, Brás et al., 2014, Karubiu et al., 2015, Yu et al., 2017, Bhakat and Söderhjelm, 2022]. In additional to the flexibility of the flap region, which was investigated in the previous studies, I looked into the side chain flipping of Trp39, which is crucial for all open flap binders. All inhibitors binding to the open flap conformation not only induce the flap opening, including a flipped Tyr75 side chain, but also induce side chain flipping of Trp39. Trp39 needs to be in an outward-pointing conformation when bound to open flap binders, as revealed by X-ray crystal structures. In order to investigate the conformational dynamics in pepsin-like aspartic proteases including the role of side chain flipping of Trp39, I analyzed the structural behavior of the flap region and surrounding residues using MD simulations and enhanced sampling methods. Three homologous proteases, renin, pepsin, and plasmepsin were studied with MD simulations, and renin was further investigated using SWISH and metadynamics simulations.

5.2   Methods

5.2.1   Clustering of X-ray crystal structures

The ligands of the renin X-ray crystal structures were clustered based on their center of mass in the binding pocket using k-means with number of clusters predefined to 3. In addition, the protein structures were clustered based on their flap backbone (Thr72 to Ser81) using the gromos method and a RMSD cutoff of 0.04 nm. The gromos method is developed by Daura et al. [1999] and follows the following procedure: The mutual RMSD between all structures is calculated and the preset RMSD cut-off determines the cluster membership. The algorithm counts the number of neighbors for each structure within the cut-off and identifies the structure with the largest number as cluster representative and assigns it and its neighbors to a cluster. For the remaining structures, this workflow is
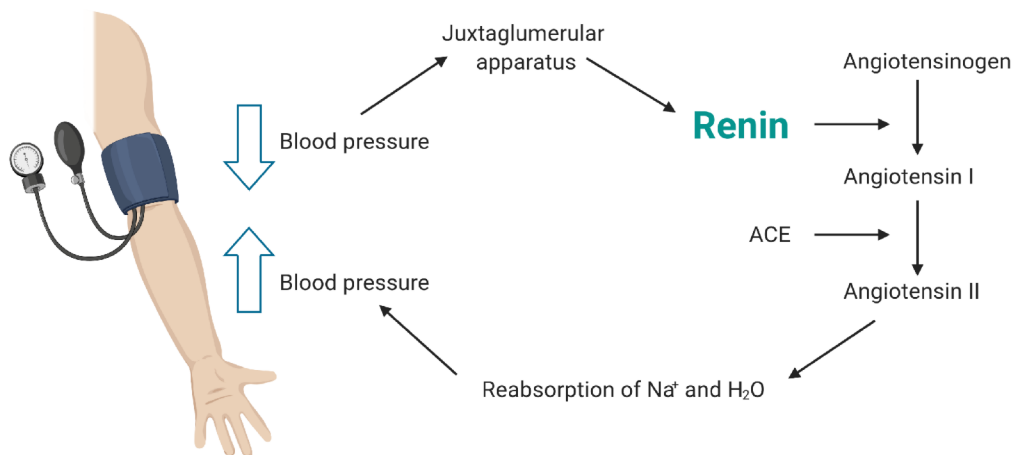
Figure 5.1: **Renin's role in the renin-angiotensin-aldosteron-system.** Renin activates the renin–angiotensin system by cleaving angiotensinogen into angiotensin 1. Created with BioRender.com.

repeated until all structures are members of a cluster. Structures without neighbors form their own clusters.

5.2.2 MD simulations

X-ray crystal structures of renin with the following PDB entries served as starting points for MD simulations: 1rne, 2bks (same coordinates as 2fs4), 2bkt, 2g26, 2g27, 3g6z, 3g70, 3g72, 3k1w, 4gj6. All ligands were removed. For 1rne, 4gj6, and 3k1w additional complex simulations were set up with the complete ligand structures for 1rne and 4gj6 and complete or fragmental ligand structure for 3k1w. As fragments either 4-bromofluorobenzene or 4-4-[3-(2-bromo-5-fluorophenoxy)propyl]phenyl-1,2,5,6-tetrahydropyridine were kept. The protein structure of 3k1w was also simulated in complex with a phenol molecule in the binding site, which was aligned on carbon atoms of 4-bromofluorobenzene. Mutated renin structures with the following point mutations in the binding site were set up. Phe112 was once mutated to Ala and once mutated to Lys in PDB entry 1rne. Two additional aspartic proteases were set up from X-ray crystal structures, PDB entries 1psn and 1lee served as starting points for pepsin and plasmepsin, respectively. All residues were simulated in their physiological protonation state, except Asp32 was protonated in all systems and His318 and His53 were doubly protonated in plasmepsin and pepsin, respectively.

Missing side chains in all set-ups were modeled using Modeller [Šali and Blundell, 1993]. Crystallographic water molecules and ions were retained. Each set-up was solvated in TIP4P-D water [Piana et al., 2015] with 150 mM NaCl each. MD simulations were carried out using Gromacs 2016 [Abraham et al., 2015] and the AMBER99SB*-ILDN-q force field [Hornak et al., 2006a, Best and Hummer, 2009, Best et al., 2012, Lindorff-Larsen et al., 2010]. Each system was energy minimized, followed by five equilibration steps, in which I gradually weakened the position restraints on heavy atoms, first in an NVT ensemble (0.25 ns) and then in an NPT ensemble (4 x 0.5 ns) using a Berendsen thermostat and barostat [Berendsen et al., 1984]. Production simulations were performed at a temperature

of 310 K and a pressure of 1 bar in an NPT ensemble using a Nosé-Hoover thermostat [Evans and Holian, 1985, Nosé, 1984] and a Parrinello-Rahman barostat [Parrinello and Rahman, 1981].

For renin (PDB entry: 1rne), 20 replicates of a mixed-solvent system were set up. 1 M benzene molecules were added to the simulation box with an interligand repulsion term to the nonbonded interactions to prevent clustering and to maintain diffusive behavior. The interligand Lennard Jones Pair was set into an effectively repulsive potential with a $\sigma = 0.8$ nm and $\varepsilon = 1.5 \cdot 10^{-3} \frac{kJ}{mol}$.

5.2.3 SWISH simulations

The SWISH simulations were performed starting from the equilibrated renin structure (PDB: 1rne) with removed ligand in Gromacs 4.6.7-plumed.2.4 using the AMBER99SB*-ILDN-q force field and the TIP4P-D water model. The replica exchange simulations were performed within a scaling factor range of 1 - 1.35 at 310 K. In total, three different SWISH simulations were set up: (1) SWISH using 8 replicas with scaling of binding site-water, (2) SWISH using 8 replicas with scaling protein-water, and (3) SWISH using 6 replicas with scaling protein-water. Parameters for the non-bonded interaction between the oxygen atoms of water molecules and carbon/sulfur atoms of the protein were calculated according to the Lorentz-Berthelot mixing rules and added to the topology. An exchange between two replicas was attempted every 2 ps. Potentials of mean force (PMF) of the unbiased replicas were obtained along the reaction coordinate $C_\alpha - C_\alpha$ distance between Asp32 and Ser76. For the (1) SWISH set-up using 8 replicas with scaling of binding site-water, the PMF was additionally calculated using the Weighted Histogram Analysis Method (WHAM) that takes information into account from all replicas.

5.2.4 Metadynamics simulations

Equilibrated renin structure (PDB: 1rne) with removed ligand structure solvated in TIP4P-D water served as starting structure for well-tempered metadynamics simulations using Gromacs 2018 and PLUMED 2 [Tribello et al., 2014]. In total, two metadynamics simulations were set up. In one set-up both the dihedral angle χ_1 of Trp39 and the $C_\alpha - C_\alpha$ distance between Asp32 and Ser76 were used as collective variables. Gaussians were deposited every 1 ps with a starting height of 0.5 $\frac{kJ}{mol}$, which gradually decreased with a bias factor of 6 and a temperature of 310 K. An upper wall for the distance variable was set to 2 nm with a force constant of 2000 $\frac{kJ}{mol}$. The widths of the Gaussians were defined on the basis of unbiased MD simulation runs and were set to 0.13 rad and 0.02 nm. In the other set-up, only the χ_1 of Trp39 served as collective variable. Gaussians were deposited every 1 ps with a starting height of 1.2 $\frac{kJ}{mol}$, a bias factor of 6 and a temperature of 310 K. The width of the Gaussians was set to 0.13 rad.

Table 5.1: Clusters of renin X-ray crystal structures based on flap coordinates

	# structures	RMSD	cluster members (PDB entries)
cluster 1	66	0.028 nm	1rne, 1bil, 1bim, 1hrn, 2g1n, 2g1o, 2g1r, 2g1s, 2g1y, 2g20, 2g21, 2g22, 2g24, 2i4q, 2iko, 2iku, 2il2, 2v0z, 2v10, 2v11, 2v12, 2v13, 2v16, 3d91, 3gw5, 3km4, 3oot, 3oqf, 3oqk, 3own, 3q3t, 3q4b, 3q5h, 3sfc, 3vsw, 3vsx, 3vuc, 3vyd, 3vye, 3vyf, 4gj7, 4gj8, 4gj9, 4gja, 4gjb, 4gjc, 4gjd, 4pyv, 4q1n, 4ryc, 4ryg, 4rz1, 4s1g, 4xx3, 4xx4, 5koq, 5kos, 5kot, 5sxn, 5sy2, 5sy3, 5sz9, 5tmg, 5tmk, 5v8v, 5vrp
cluster 2	9	0.032 nm	3g6z, 3g70, 3g72, 3k1w, 3o9l, 3oad, 3oag, 4gj6, 5vpm
cluster 3	2	0.000 nm	2bks, 2fs4
cluster 4	1		2bkt
cluster 5	1		2g26 (chain B)
cluster 6	1		2g27
cluster 7	1		4gj5

5.3 Results and discussion

5.3.1 Renin structures show different flap conformation

Renin can adopt different flap conformations from closed to open states. All X-ray crystal structures from the PDB without a ligand (PDB entries: 1bbs, 2ren, 2x0b) show the closed flap conformation. X-ray crystal structures with bound ligands show either closed or open flap conformations, depending on the ligand binding mode. For all complex X-ray crystal structures in the PDB, I clustered the ligands based on their center of mass in the binding pocket using k-means with number of clusters predefined to 3. In Fig. 5.2a,b the complex X-ray structures are shown with the three clusters of ligands (shown in red, yellow, and green). Additionally, I clustered the protein structures based on their flap backbone (Thr72 to Ser81) using the gromos method and a RMSD cutoff of 0.04 nm. Using all X-ray crystal structures of renin, except PDB entry 5t4s (due to missing flap coordinates), this procedure outputted seven clusters as listed in Table 5.1.

Looking at the clusters, closed flap structures belong to cluster 1 and open flap structures are present in cluster 2. The remaining clusters are one-membered clusters (PDB entries: 2bks and 2fs4 are identical). In order to assign them to either the closed or open flap clusters, I calculated the RMSD values between the remaining clusters and both cluster representative structures of cluster 1 and 2. The representative structures for cluster 1 (closed flap) and cluster 2 (open flap) are 5sxn and 3oad, respectively. The RMSD values are listed in Table 5.2. Due to lower RMSD values to the corresponding cluster and visual inspection, 2bks/2fs4, 2bkt, 2g27 are assigned to the open flap and 4gj5 to the closed flap structures. 2g26 (chain B) fits more into the open flap structures, but since the flap is curled (Fig. 5.2f) the conformation is referred to as distorted. The clustering procedure assigned the PDB entries 5vpm and 4gj6 to an open flap conformation. However, both χ_1 dihedral angles of Trp39 and Tyr75 of 5vpm and 4gj6 show an inward-pointing conformation

Table 5.2: RMSD to cluster 1 and 2 for cluster 3, 4, 5, 6, 7

	PDB entries	RMSD to cluster 1 (closed flap)	RMSD to cluster 2 (open flap)
cluster 3	2bks, 2fs4	0.778 nm	0.5587 nm
cluster 4	2bkt	0.7539 nm	0.5055 nm
cluster 5	2g26 (chain B)	0.7792 nm	0.8357 nm
cluster 6	2g27	0.89790 nm	0.7594 nm
cluster 7	4gj5	0.47166 nm	0.9026 nm

Table 5.3: Structures used to generate Fig. 5.2

PDB entries of closed flap conformation	PDB entries of open flap conformation
1rne, 1bil, 1bim, 1hrn, 2g1n, 2g1o, 2g1r, 2g1s, 2g1y, 2g20, 2g21, 2g22, 2g24, 2i4q, 2iko, 2iku, 2il2, 2v0z, 2v10, 2v11, 2v12, 2v13, 2v16, 3d91, 3gw5, 3km4, 3oot, 3oqf, 3oqk, 3own, 3q3t, 3q4b, 3q5h, 3sfc, 3vsw, 3vsx, 3vuc, 3vyd, 3vye, 3vyf, 4gj5, 4gj7, 4gj8, 4gj9, 4gja, 4gjb, 4gjc, 4gjd, 4pyv, 4q1n, 4ryc, 4ryg, 4rz1, 4s1g, 4xx3, 4xx4, 5koq, 5kos, 5kot, 5sxn, 5sy2, 5sy3, 5sz9, 5t4s, 5tmg, 5tmk, 5v8v, 5vrp	2bks, 2bkt, 2fs4, 2g27, 3g6z, 3g70, 3g72, 3k1w, 3o9l, 3oad, 3oag

(Fig. 5.2f). Hence, they are labeled as semi-open in the following. A final list of all open and closed flap structures can be found in Table 5.3. All X-ray crystal structures of renin are displayed in Fig. 5.2.

As depicted in Fig. 5.2e all open flap conformations display a similar ligand binding mode. All ligands inducing the flap opening are members of the same cluster (green) and are located deeper in the binding pocket than ligands from the other two clusters. The ligands penetrate the binding pocket, leading to further pocket opening. As shown in Fig. 5.2f the Trp39 and Tyr75 side chains are flipped in the open flap conformations. Ligand arms displace the side chain of Trp39 and wedge in between Trp39 and Phe112.

5.3.2 Flap in renin closed in unbiased MD simulations

To investigate the flap dynamics, I set up a long MD simulation with the closed flap as starting conformation (PDB entry: 1rne) and eight short MD simulations with the open flap as starting conformations (PDB entries: 2bks, 2bkt, 2g26, 2g27, 3g6z, 3g70, 3g72, 3k1w) with removed ligand structures (Fig. 5.3a,b). The flap remained closed over the whole course of the trajectory of 4 µs or, if started from an open state, closes within a timescale of 500 ns (Fig. 5.3c,d).

The Trp39 side chain flipped in 50% of the MD simulations starting from the open flap conformations towards the closed conformation, but didn't change its conformation in the 4 µs long MD simulation starting from the closed conformation. The MD simulations of PBD entries: 2bkt, 2g26 and 2g27 showed a fast Trp39 side chain flipping at the start of the simulations. The Trp39 side chain in MD simulation from PDB entry 3g72 changed its open conformation towards the closed conformation at ≈400 ns. In the four remaining
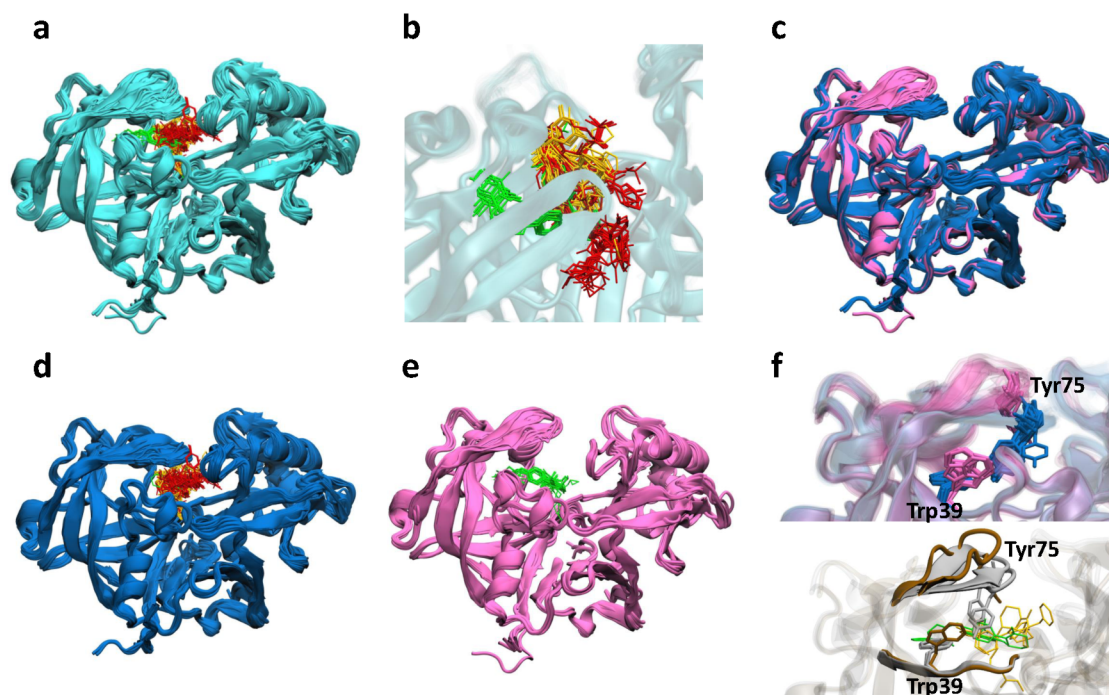
Figure 5.2: **Superimposed X-ray crystal structures of renin. a,** Superimposed X-ray crystal structures of renin in cartoon, with ligands shown with atomic detail. Color of ligands denote three clusters based on center of mass. **b,** Same structures as in a) shown from above with transparent protein. **c,** Open and closed flap X-ray crystal structures shown in magenta and blue, respectively. **d,** closed flap X-ray crystal structures with corresponding ligands. **e,** open flap X-ray crystal structures with corresponding ligands. **e,** Top: Open and closed flap X-ray crystal structures with Trp39 and Tyr75 depicted with atomic detail shown in magenta and blue, respectively. Bottom: Semi-open flap X-ray crystal structures (PDB entry: 2g26 (chain B) in brown and PDB entries: 5vpm, 4gj6 in gray) with Trp39 and Tyr75 depicted with atomic detail and with corresponding ligands.
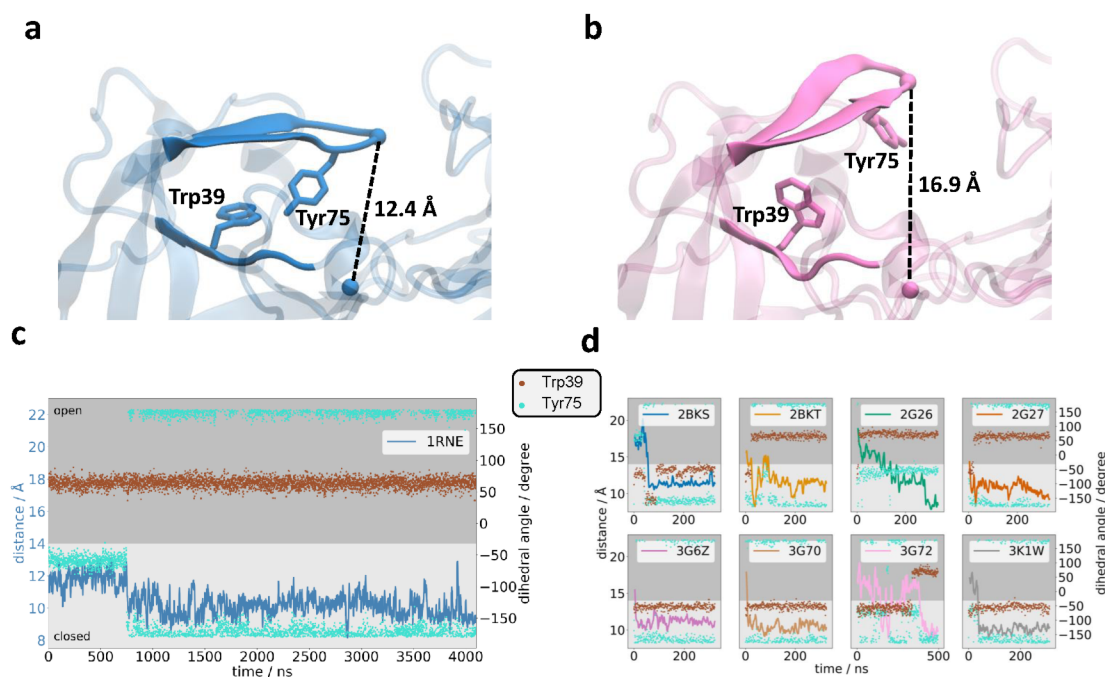
Figure 5.3: **MD simulations of renin. a,** X-ray crystal structure of PDB entry 1rne shows typical closed flap conformation. Distance denotes $C_\alpha - C_\alpha$ distance between Asp32 and Ser76. **b,** X-ray crystal structure of PDB entry 3g72 shows typical open flap conformation. Distance denotes $C_\alpha - C_\alpha$ distance between Asp32 and Ser76. **c,** Flap remained closed in unbiased MD simulation of PDB entry 1rne. $C_\alpha - C_\alpha$ distance between Asp32 and Ser76 (blue), χ_1 angle of Trp39 (brown) and χ_1 angle of Tyr75 (cyan) are shown as functions of time. **d,** Flap closed in unbiased MD simulations of open flap conformations as starting structures (PDB entries: 2bks, 2bkt, 2g26, 2g27, 3g6z, 3g70, 3g72, 3k1w). $C_\alpha - C_\alpha$ distance between Asp32 and Ser76 (blue), χ_1 angle of Trp39 (brown) and χ_1 angle of Tyr75 (cyan) are shown as functions of time.

MD simulations starting from the open conformation, Trp39 either remained in its starting conformation (PBD entries: 3g6z, 3g70, and 3k1w) or changed its conformation transiently to a new conformation with a χ_1 angle of -150°.

The Tyr75 side chain showed some flexibility in all simulations and didn't change its conformation only in three runs with the PDB entries 2g27, 3g6z, and 3g70 as starting structures. In the long MD simulation starting from the closed conformation, the Tyr75 side chain flipped at $\approx$750 ns, which led to further flap closing. In all simulations, the combination of χ_1 angle of Trp39 at 70° and the χ_1 angle of Tyr75 at (-)170° showed a closed flap conformation.

Side chain flipping of Trp39 and Tyr75 seems to be correlated with the $C_\alpha - C_\alpha$ distance between Asp32 and Ser76, but not consistently. Side chain flipping can lead to (further) flap closing as observed in the MD simulations from PDB entries 1rne, 2bks, 2bkt, 2g27 3g72 and 3k1w. The flap from PDB entries 3g6z and 3g70 closed spontaneously independent of side chain flipping, probably due to hydrophobic interactions.

Starting from the closed conformation, the long unbiased MD simulation was not able to sample flap opening and the full pocket opening with flipped Trp39 and Tyr75. To enhance the sampling, I used simulations with increased water-protein interaction and mixed-solvent simulations, as described in Chapter 5.3.4.

5.3.3 Minimal truncated form of open flap binder needs to have interactions with catalytic Asp to stabilize the flap in open form

To assess the complex stability, I set up simulations of the closed flap complex (PDB entry: 1rne) and one open flap complex (PDB entry: 3k1w) in presence of the ligand for each structure. Both ligand structures remained stably bound, displaying the same binding mode as observed in the X-ray crystal structures (Fig. 5.4a,b,f). Starting from the closed flap conformation, the flap showed some flexibility and Tyr75 side chain flipping (Fig. 5.4a), while the open flap binder BFX locked the flap coordinates into the open form (Fig. 5.4b). The conformation of Trp39 remained unchanged in both simulations.

To understand which fragment of BFX is essential for locking the flap in the open flap form, three additional simulations were set up: (1) large fragment of BFX (4-4-[3-(2-bromo-5-fluorophenoxy)propyl]phenyl-1,2,5,6-tetrahydropyridine), (2) small fragment of BFX (4-bromofluorobenzene) and (3) a phenol molecule aligned on carbon atoms of 4-bromofluorobenzene of BFX. Only the large fragment of BFX was able to lock the flap in an open flap form, similar to BFX (Fig. 5.4c). The small fragment, 4-bromofluorobenzene, remained bound in the hydrophobic pocket of Trp39 and Phe112, but showed some flexibility (Fig. 5.4d,f). Bound 4-bromofluorobenzene did not rigidify the open form, and the flap closed during the simulation (Fig. 5.4d). Replacing 4-bromofluorobenzene with phenol resulted in phenol leaving the binding pocket and flap closing (Fig. 5.4e,f). Hence, the electrostatic interactions with the catalytic Asp and the hydrophobic interactions with Trp39 and Phe112 are necessary for a compound to stabilize the flap in an open form.

5.3.4 Scaling water-protein interaction and mixed-solvent simulations led to more flap dynamics, but did not sample fully open pocket with flipped Trp39

Gervasio et al. established a Hamiltonian replica-exchange method called SWISH (sampling water interfaces through scaled Hamiltonians) that enhances the sampling of hydrophobic pocket openings by increasing the interactions between water molecules and protein atoms. I set up three SWISH protocols of renin: (1) scaling only the interactions with residues of the binding pocket using 8 replicas, (2) scaling interactions with all protein residues using 8 replicas, and (3) scaling interactions with all protein residues using 6 replicas. For all procedures, I used the closed flap conformation without ligand structure (PDB entry: 1rne) as starting structure. Indeed, flap openings were observed (Fig. 5.5) in contrast to the 4 μs unbiased simulation, where no flap opening took place. Interestingly, sampling of the distinctive conformational states worked best in the set-up, where only the interaction of the binding site were scaled (Fig. 5.5a). The simulations found three conformational states, very closed (8 Å), closed (12 Å) and open (16 Å). The two set-ups with scaling all protein residues also sampled open flap forms, but with the open flap form scarcely populated. In the set-up with 6 replicas, the replica exchange with replica 5 and 6 was stopped within the first 100 ns, thus in Fig. 5.5f only the replicas 1 to 4 are plotted. The energy barrier between the closed and open form was similar between all set-ups and was estimated to ≈3 kT. In all three SWISH simulations, side chain flipping of Tyr75 was observed in all
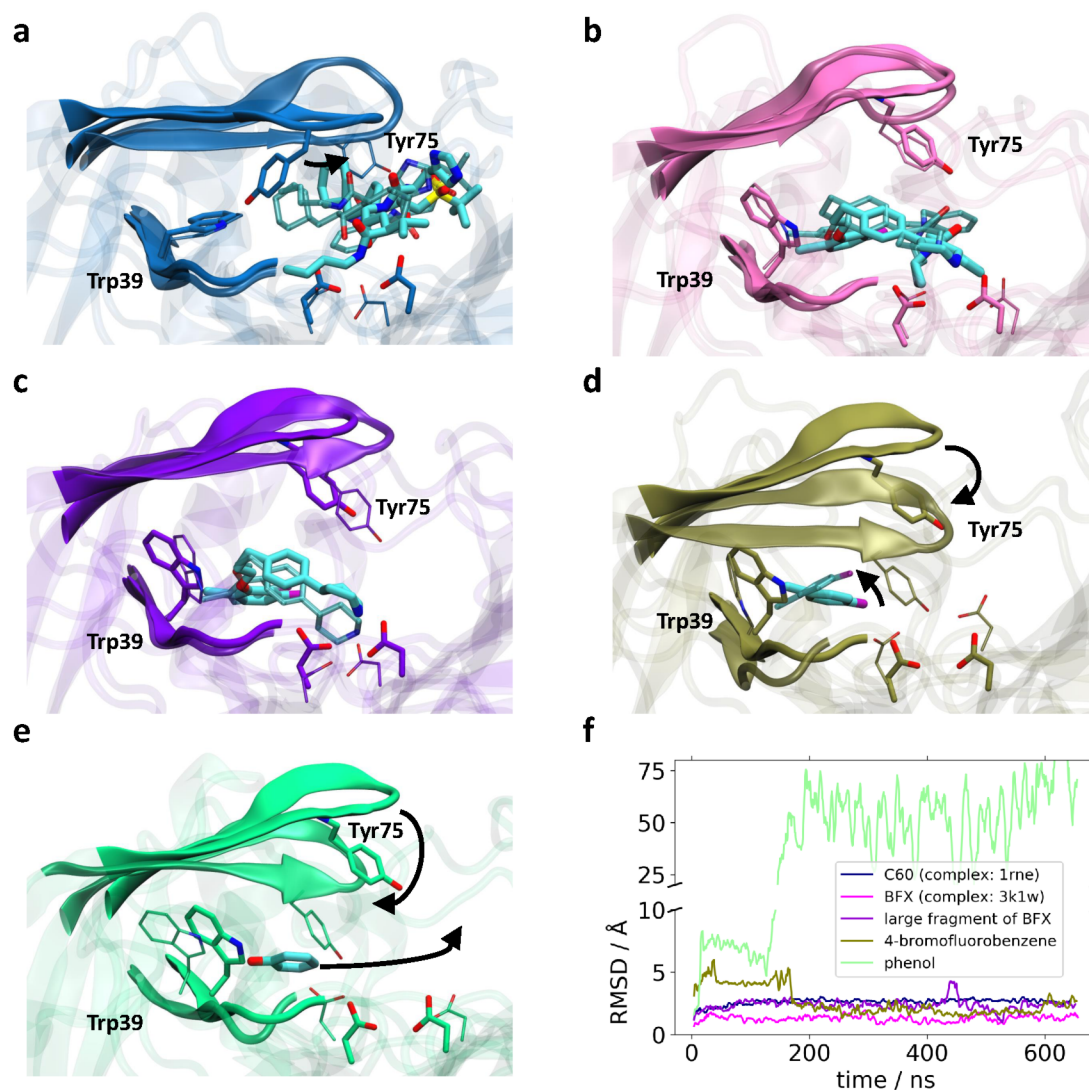
Figure 5.4: **MD simulations of ligand-bound renin systems. a-e** Superposition of first (protein displayed in opaque cartoon, residues and ligand in big stick representation) and last MD frame (protein displayed in shiny and transparent cartoon, residues and ligand in thin stick representation). Black arrows denote changes during simulations. **a,** complex of PDB entry 1rne, **b,** complex of PDB entry 3k1w. **c,** complex of PDB entry 3k1w with truncated ligand (4-4-[3-(2-bromo-5-fluorophenoxy)propyl]phenyl-1,2,5,6-tetrahydropyridine). **d,** complex of PDB entry 3k1w with truncated ligand (4-bromofluorobenzene). **e,** complex of PDB entry 3k1w with phenol as ligand. **f,** heavy-atom RMSD of ligands as function of time with respect to each starting structure.

replicas, but Trp39 always remained in its inward-pointing conformation ($\chi_1 = \approx 70°$).

To enhance sampling further, I set up 20 mixed-solvent simulations of renin, starting from the closed conformation with 1 M benzene molecules (Fig. 5.6a). In 20 x 500 ns trajectories, increased flap dynamics was observed and the side chain of Tyr75 adopted various conformations. However, the side chain of Trp39 remained unchanged within the accumulated 10 µs (Fig. 5.6b).

SWISH and mixed-solvent simulations starting from the closed conformation led to more flap dynamics and Tyr75 side chain flipping compared to the unbiased MD simulation of 4 µs, but did not sample the full pocket opening with flipped Trp39.

5.3.5 Point mutations of Phe112 in renin did not result in flipped Trp39

In the closed flap conformation, as seen in X-ray crystal structures, Trp39 has hydrophobic interactions with Phe112. More hydrophobic side chains (Leu71, Leu73 and Val80) are located at the bottom of the flap. In wild-type renin simulations from PDB entry 1rne, I solely observed water molecules entering below the flap region. To understand the Trp39 side chain dynamics better, I computationally mutated Phe112 once to Ala, and once to Lys, to reduce the hydrophobic interactions. In both MD simulations with the point mutations, the number of water molecules entering into the pocket increased. However, the conformation of the side chain of Trp39 remained rigid, as observed in the WT simulations (Fig. 5.7d).

5.3.6 Pepsin and plasmepsin showed similar behavior as renin

Renin is a member of the aspartic protease family, which also includes several other proteins like gastricsin, pepsin, cathepsin D and BACE (human), plasmepsin (protozoan) and HIV-1 protease (viral). To understand whether the Trp39 side chain conformation is similarly stable in related proteins, I selected two drug targets similar to renin for MD simulation: pepsin and plasmepsin.

Both of them have a Trp at 39 position. Plasmepsin (Fig. 5.8b) has a Tyr at 112 position, and pepsin (Fig. 5.8c) has a Val instead of Phe112. In both trajectories of 2 µs each, the flap showed some flexibility, but Trp39 remained in its starting conformation. Hence, these simulations indicate that the rigidity of Trp39 is not only present in renin, but also in pepsin and plasmepsin, and a characteristic feature of the pepsin-like protease family.

5.3.7 Well-tempered metadynamics simulation indicated a large free energy barrier of Trp39 side chain flipping

Even with a lot of computational effort, 4 µs of unbiased MD simulations, 10 µs of mixed-solvend MD simulations, and SWISH simulations, Trp39 side chain flipping starting from the closed flap conformation was not observed. In contrast, starting from the open flap conformation with outward-pointing Trp39 side chain and without ligand structure could result in Trp39 swinging back to its inward-pointing conformation. Thus, the energy barrier for the
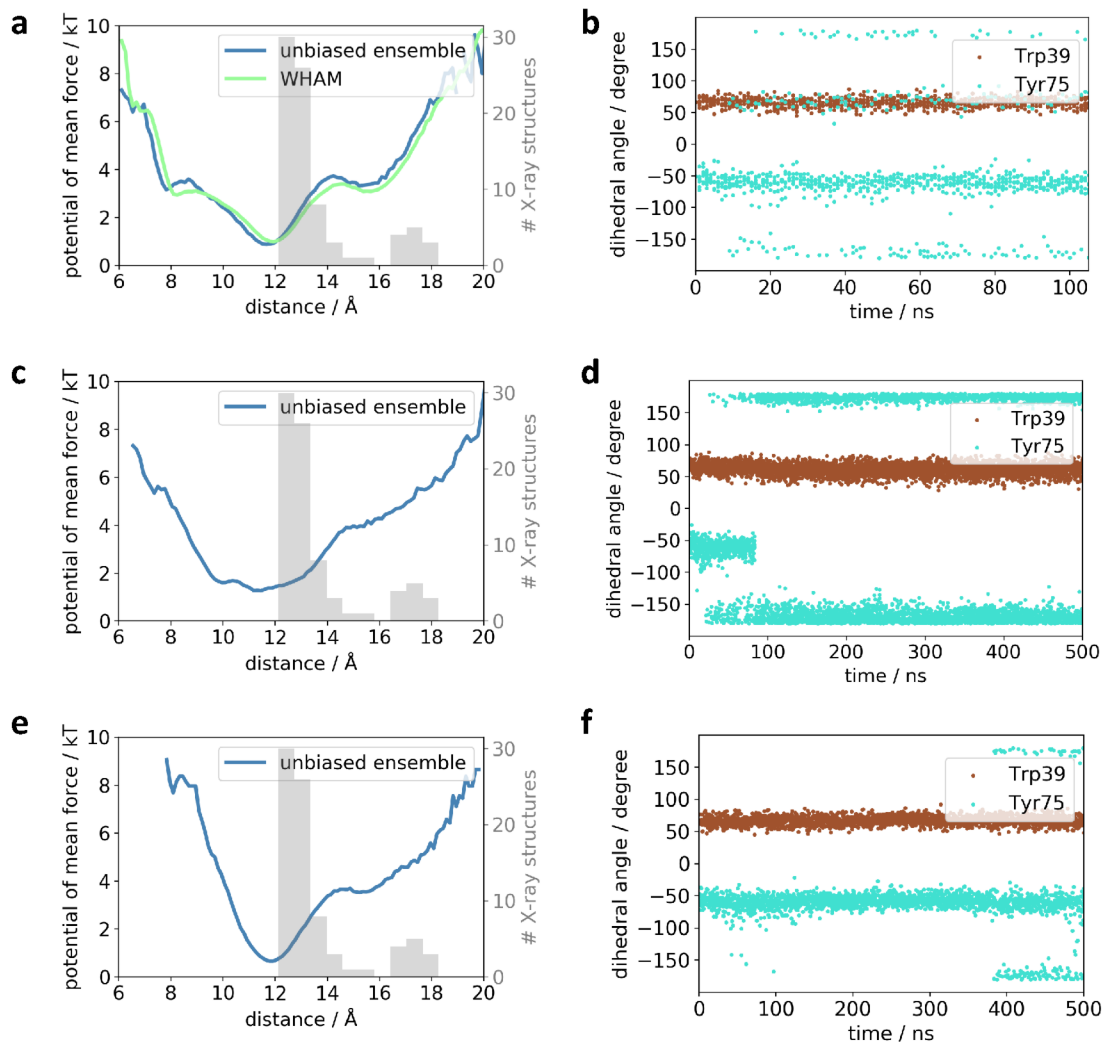
Figure 5.5: **SWISH simulations of renin. top: SWISH with 8 replicas and scaling of binding site-water. a,** free energy profile of $C_\alpha - C_\alpha$ distance between Asp32 and Ser76 for unbiased ensemble (lowest replica, blue) and for all ensembles calculated with WHAM (green) with histogram of values from X-ray structures. **b,** dihedral angles $\c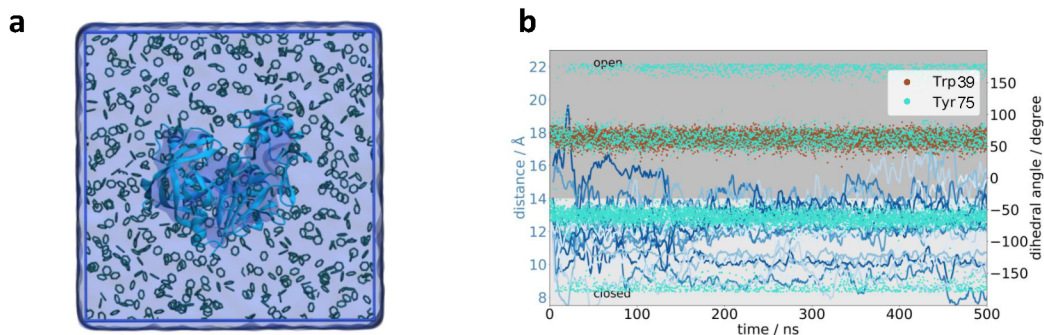hi_1$ of Trp39 (brown) and Tyr75 (cyan) as function of time. **middle: SWISH with 8 replicas and scaling protein-water. c,** free energy profile of $C_\alpha - C_\alpha$ distance between Asp32 and Ser76 for unbiased ensemble (lowest replica) with histogram of values from X-ray structures. **d,** dihedral angles χ_1 of Trp39 (brown) and Tyr75 (cyan) as function of time. **bottom: SWISH with 6 replicas and scaling protein-water. e,** free energy profile of $C_\alpha - C_\alpha$ distance between Asp32 and Ser76 for unbiased ensemble (lowest replica) with histogram of values from X-ray structures. **f,** dihedral angles χ_1 of Trp39 (brown) and Tyr75 (cyan) as function of time.

Figure 5.6: **Mixed-solvent simulations of renin. a,** Simulation box of mixed-solvent set-up with renin and 1 M benzene molecules shown in cartoon and with atomic detail, respectively. **b,** Flap opened in mixed-solvent simulations of closed flap conformation as starting structure (PDB entry: 1rne), but Trp39 remained in its starting conformation. $C_\alpha - C_\alpha$ distances between Asp32 and Ser76 are shown in line plots in different shades of blue for each replica, χ_1 angle of Trp39 (brown) and χ_1 angle of Tyr75 (cyan) are shown as functions of time.
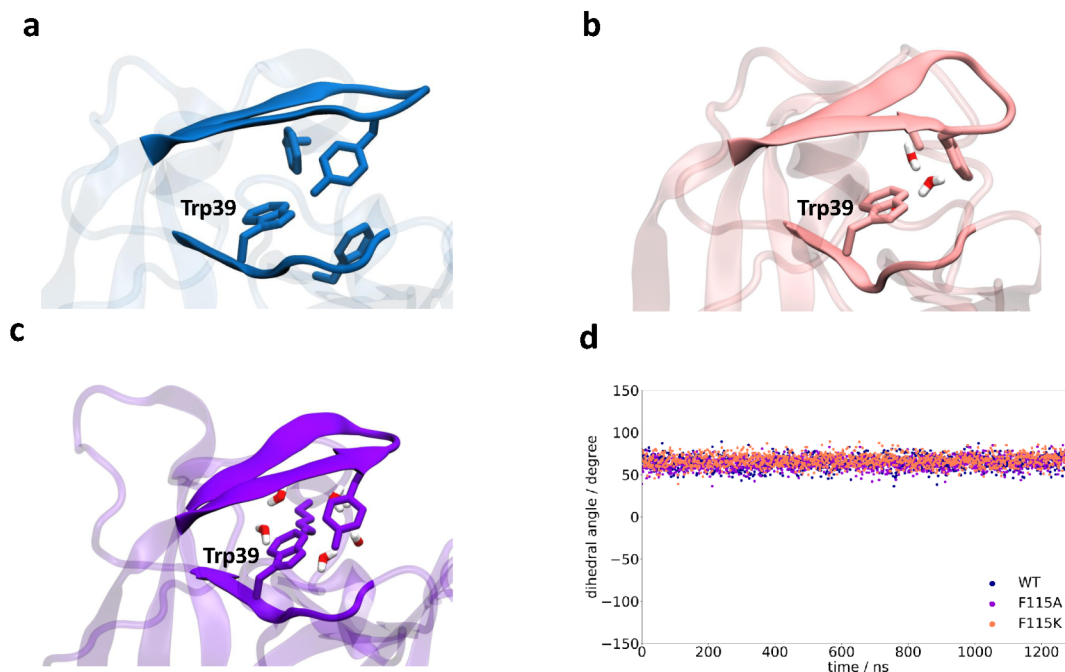
Figure 5.7: **Mutations of Phe112. a,** X-ray crystal structure of wild-type renin in cartoon with Trp39, Phe112 and Tyr75 shown with atomic detail. **b,** last MD frame of renin with F115A mutation in cartoon with Trp39, Ala115 and Tyr75 shown with atomic detail. **c,** last MD frame of renin with F115K mutation in cartoon with Trp39, Lys115 and Tyr75 shown with atomic detail. **d,** χ_1 angle of Trp39 are shown in blue (wild-type), purple (F115A) and light pink (F115K) as functions of time.
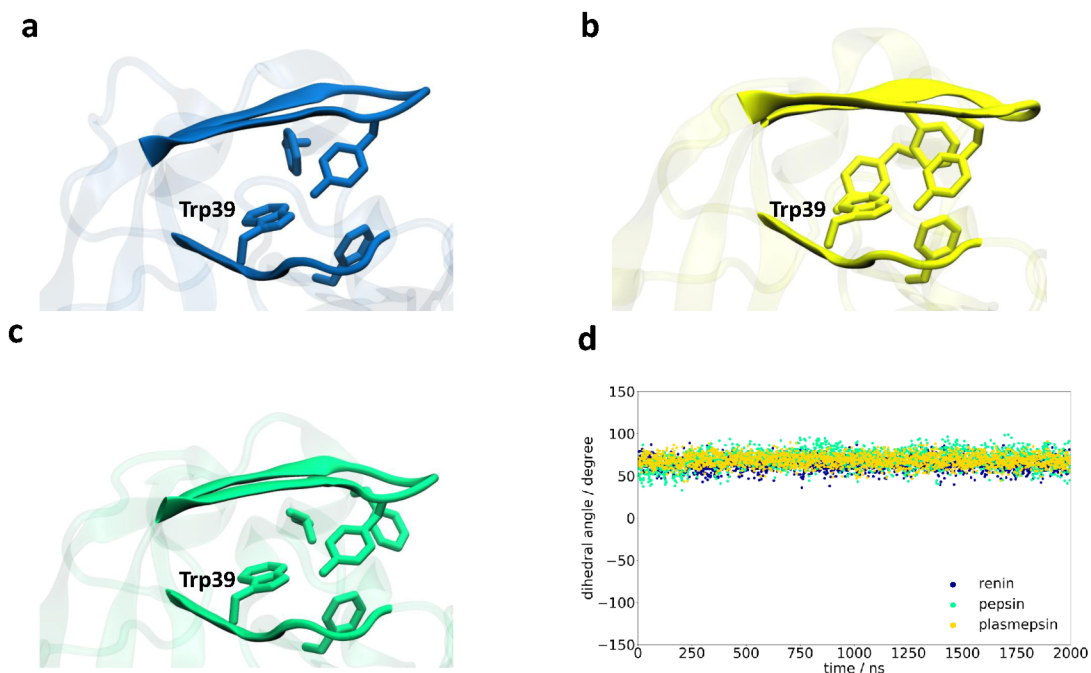
Figure 5.8: **Pepsin and plasmepsin. a,** X-ray crystal structure of renin in cartoon with Trp39, Phe112 and Tyr75 shown with atomic detail. **b,** X-ray crystal structure of plasmepsin in cartoon with Trp39, Phe112 and Tyr75 shown with atomic detail. **c,** X-ray crystal structure of pepsin in cartoon with Trp39, Phe112 and Tyr75 shown with atomic detail. **d,** χ_1 angle of Trp39 are shown in blue (renin), green (pepsin) and yellow (plasmepsin) as functions of time.

Trp39 outwards side chain flipping is expected to be large. In order to gain insight into the energy landscape underlying the conformational change, two well-tempered metadynamics simulations were performed. In one set-up both the dihedral angle χ_1 of Trp39 and the $C_\alpha - C_\alpha$ distance between Asp32 and Ser76 were biased, and in the other one only the dihedral angle χ_1 of Trp39 was used as collective variable. In the first set-up, one forward and one backward transition of Trp39 was obtained within 800 ns (Fig. 5.9a,b). The flap, which was also biased, changed its conformation often. Biasing only the Trp39 χ_1 dihedral angle led to two forward and to two backward transitions within 2 μs (Fig. 5.10a,d). Interestingly, the second transition took place without deposition of an additional Gaussian (Fig. 5.10e). This indicates that the simulation from this part onwards was run on a biased potential accumulated through the initial deposited Gaussians. For each χ_1 dihedral transition, the flap opened simultaneously, suggesting that Trp39 side chain flipping requires an open flap (Fig. 5.10d). Both metadynamics simulations yielded a free energy barrier of $\approx 35 \frac{kJ}{mol}$ between the flipped-in and the flipped-out Trp39 side chain. The large energy barrier for the conformational change explains the sampling difficulties with previous methods.

5.4   Summary and conclusion

The flap dynamics in pepsin-like aspartic proteases is important for substrate entrance and catalytic activity. The majority of aspartic proteases possess the conserved amino acids Trp39 and Tyr75 close to the active site, which is covered by the flap. Here, using unbiased
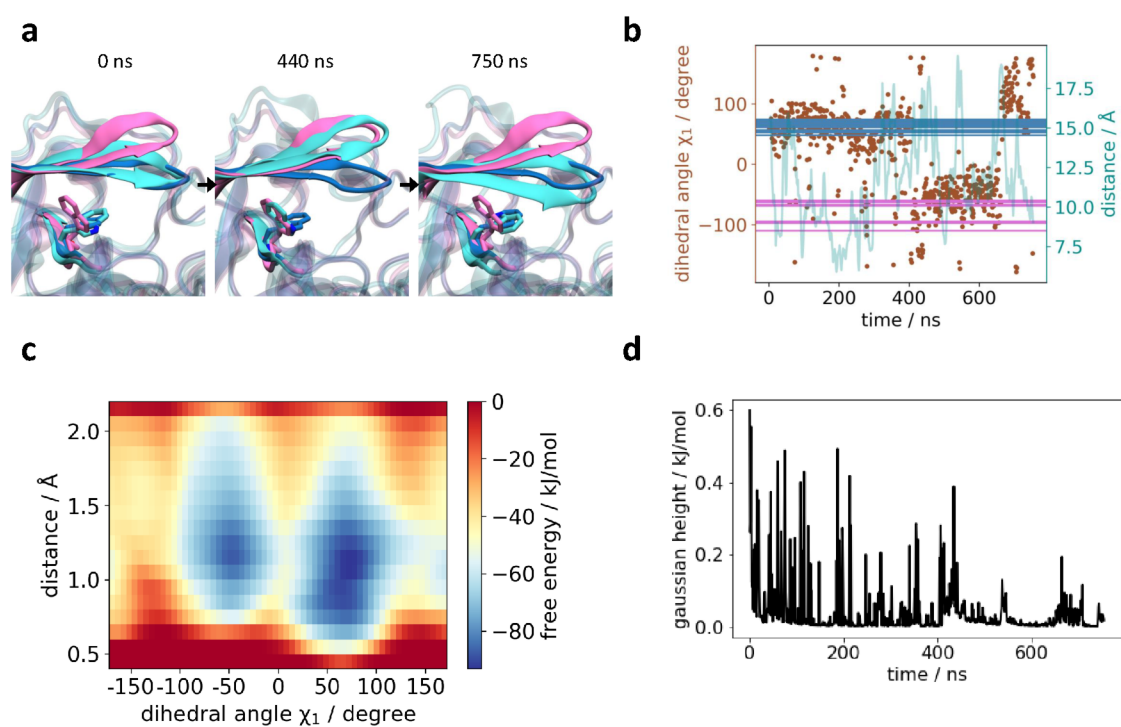
Figure 5.9: **Metadynamics simulation of renin with biasing χ_1 of Trp39 and flap distance.a,** Snapshots of superpositions of metadynamics simulation frames in cyan with X-ray crystal structures of closed flap (blue) and open flap structures (magenta). **b,** $C_\alpha - C_\alpha$ distance between Asp32 and Ser76 and χ_1 dihedral angle as function of time, blue and magenta lines denote χ_1 dihedral angle values of X-ray crystal structures. **c,** 2D free energy profile of χ_1 dihedral angle and $C_\alpha - C_\alpha$ distance between Asp32 and Ser76. **d,** Gaussian height as a function of time.
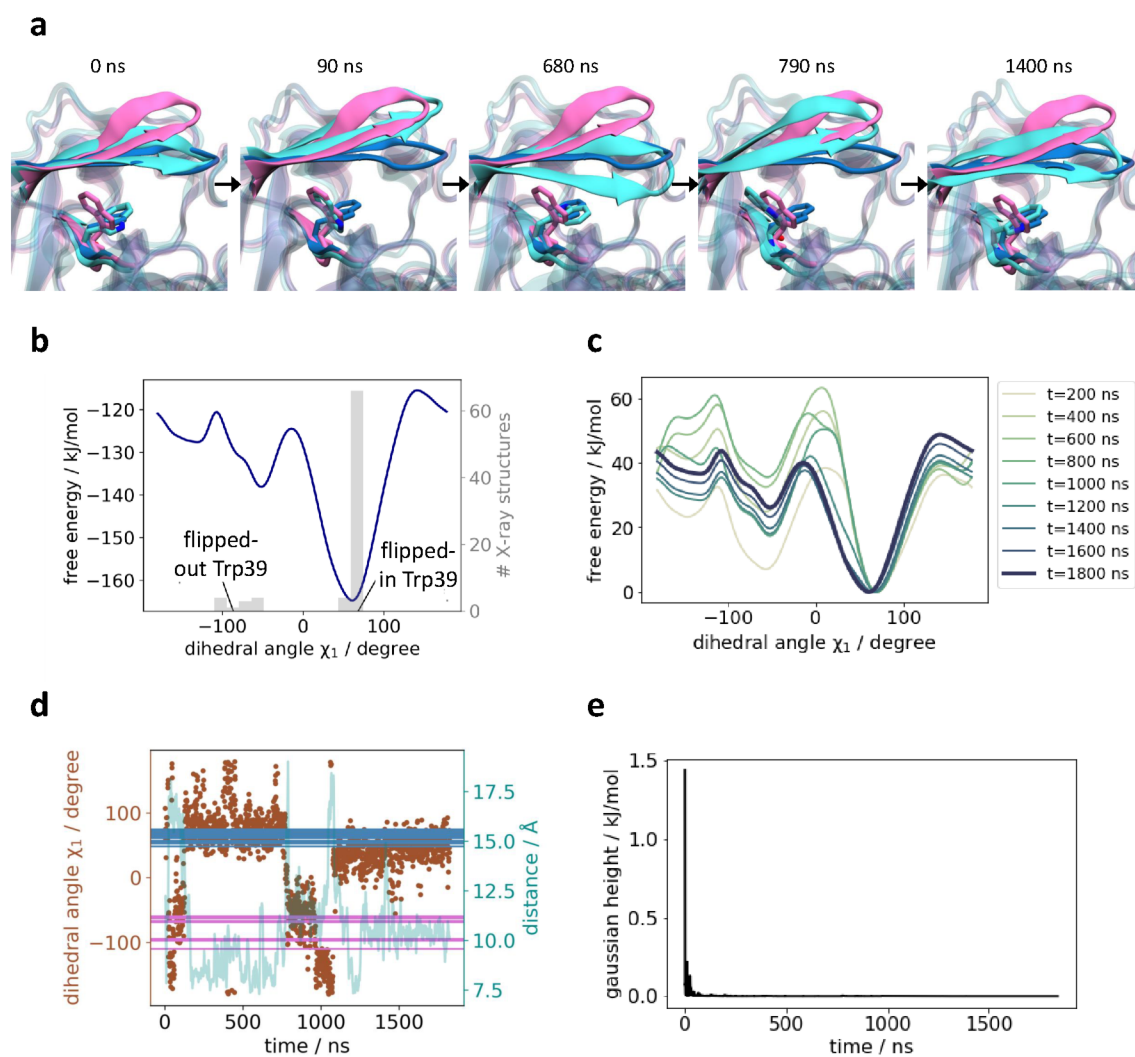
Figure 5.10: **Metadynamics simulation of renin with biasing χ_1 of Trp39. a,** Snapshots of superpositions of metadynamics simulation frames in cyan with X-ray crystal structures of closed flap (blue) and open flap structures (magenta). **b,** free energy profile of χ_1 dihedral angle with histogram of values from X-ray structures. **c,** convergence of the free energy profile of χ_1 dihedral angle. **d,** $C_\alpha - C_\alpha$ distance between Asp32 and Ser76 and χ_1 dihedral angle as function of time, blue and magenta lines denote χ_1 dihedral angle values of X-ray crystal structures. **e,** Gaussian height as a function of time.

and biased MD simulations, I investigated the plasticity of the flap and surrounding residues in aspartic proteases, with a focus on renin. I observed that the flap can possess closed and open conformations including Tyr75 side chain flipping, which is in good agreement with previous studies [Xu et al., 2012, Brás et al., 2014, Karubiu et al., 2015, Yu et al., 2017]. Very recently Bhakat and Söderhjelm [2022] studied the flap dynamics of pepsin-like proteases with great detail. They found that the flap of the free enzyme can transition between various states, closed and open forms, and that the side chain flipping of Tyr75 potentially plays a decisive role for the flexibility of the flap. In their two studied systems, plasmepsin-II and BACE-1, computational mutation of Tyr75 to Ala resulted in complete flap collapse. Interestingly, previous experimental studies showed that the same mutation led to a loss of activity [NAsIR et al., 1999]. Here, I observed similar flexibility of the flap of the wild-type apo form, including Tyr75 side chain flipping. In all 20 MD runs with 1 M benzene molecules, the flap changes its conformation between closed and open forms. However, in the 4 μs unbiased MD simulation run without 1 M benzene molecules, the flap remained closed, probably due to limited sampling time.

In previous studies, only little attention was given to the dynamics of the Trp39 residue close to the flap. It is known that Trp39 can form a H-bond with Tyr75, stabilizing the closed flap conformation [Bhakat and Söderhjelm, 2022]. Comparing all available X-ray crystal structures reveals that all inhibitors binding to the open flap conformation also displace the Trp39 side chain. Hence, even though flexible flap dynamics can be sampled with simulations, the transition to a fully open hydrophobic subpocket, as observed in X-ray crystal structures, is only achieved with an outward flipped Trp39 side chain and has not been observed until now. Here, I could sample the Trp39 side chain flipping, but only via applying specific biasing. I did not observe Trp39 side chain flipping in renin, pepsin, and plasmepsin using MD, mixed-solvent MD, and SWISH simulations. Only well-tempered metadynamics runs biasing the Trp39 dihedral angle χ_1 led to the sampling of outward-flipped Trp39 side chain in renin as observed in all open flap X-ray crystal structures. This suggests that even long simulations, and biased simulations where no specific bias on the conformational change is applied, fall short, when dealing with transient, high-energy pockets. The well-tempered metadynamics simulations showed that a large energy barrier of 35 $\frac{kJ}{mol}$ limits the flexibility of Trp39 side chain, and the side chain of Trp39 is commonly populated in an inward pointing position. Certain small molecules can displace the side chain of Trp39 to the flipped-out conformation. Hence, our simulations indicate that small molecules are capable of inducing major conformational changes, up to 35 $\frac{kJ}{mol}$, yet achieve high-affinity binding. This implies that these ligands potentially have an exceptionally long on-time, and simultaneously a slow off rate. This can be exploited for further drug design studies.

Most of the aspartyl proteases possess structural and sequence similarity, indicating that understanding the conformational dynamics associated with a subset of these enzymes allows for understanding other homologous enzymes. Hence, I postulate that the high energy barrier of side chain flipping of Trp39 is also present in other enzymes of the pepsin-like protease family.

Cryptic pockets in NPC2, p38α, ricin, and Eg5

6.1 Introduction

How cryptic pockets are formed is not very well understood. Even though ligands seem to be required for the opening of the cryptic pockets, it is not clear whether the binding process follows an induced-fit, conformational selection, or a "mixed" mechanism [Csermely et al., 2010]. Due to this lack of knowledge and their hidden nature, cryptic sites are difficult to find by both experimental and computational methods. Physics-based approaches, such as MD simulations, are limited by the possible sampling time and machine learning approaches lack the availability of large and high-quality experimental data sets.

The CryptoSite data set by Cimermancic et al. [2016] consists of a representative set of 93 unbound and ligand-bound pairs with cryptic sites, including 4 proteins with previously known cryptic sites from the literature (exportin-1, TEM1 β-lactamase, IL-2, and Bcl-X). This data set is visualized in the Appendix. FVIIa and thrombin, the trypsin-like proteases studied in Chapter 4, are part of the data set, however Cimermancic et al. [2016] selected them based on older X-ray crystal structures not displaying the major structural changes investigated in this Thesis. In addition to FVIIa and renin, I investigated protein conformational changes upon ligand binding in four other systems selected from the CryptoSite data set: NPC2, p38α, ricin, and Eg5. In all four systems, it is known from X-ray crystal structures that ligand binding can cause conformational changes. Our aim is to sample the conformational space made accessible upon binding of the ligands, yet without using the specific ligand structures or prior knowledge on the location and features of the pockets. We compared three different simulation methods, which can be beneficial for identifying altered protein structures that can be targeted. I used these methods to probe and induce the opening of cryptic pockets starting from the occluded apo structures: MD simulations, mixed-solvent MD simulations with 1 M benzene molecules, and an enhanced sampling method called SWISH. This allowed us to investigate the opening mechanisms of the selected subset. In this Chapter, I summarize the results that were generated together with Simon Tiede, a former Bachelor Thesis student working with me [Tiede, 2021].

6.2 Methods

6.2.1 MD simulations

The coordinates of all proteins in their apo conformation were taken from the PDB database (table 6.1) and simulated in their physiological protonation state. For p38α, the missing loop between residues 170 and 185 was modeled using the additional X-ray crystal structure

from PDB entry: 2baj [Sullivan et al., 2005] after backbone alignment. All ligand structures were removed, except ANP^{4-} in Eg5, which was parameterized with GAFF [Wang et al., 2004]. Missing side chains were modeled using MODELLER [Šali and Blundell, 1993] and all proteins were solvated in TIP3P water [Jorgensen et al., 1983] with 150 mM NaCl.

protein	PDB entry (apo)	PDB entry (ligand-bound)
NPC2	1nep [Friedland et al., 2003]	2hka (chain C) [Xu et al., 2007]
p38α	4e5b [Tzarum et al., 2012]	3hl7 [Xing et al., 2009]
ricin	1rtc [Mlsna et al., 1993]	1br6 [Yan et al., 1997]
Eg5	3hqd [Parke et al., 2010]	1q0b (chain B) [Yan et al., 2004]

Table 6.1: PDB entries for investigated proteins

MD simulations were carried out using the AMBER99SB*-ILDN-q force field [Hornak et al., 2006a, Best and Hummer, 2009, Best et al., 2012, Lindorff-Larsen et al., 2010]. Each system was energy minimized, followed by five equilibration steps, in which the position restraints on heavy atoms were gradually weakened, first in an NVT ensemble (0.1 ns) and then in an NPT ensemble (4 x 0.5 ns) using a Berendsen thermostat and barostat [Berendsen et al., 1984]. Production simulations were run at a temperature of 310 K and a pressure of 1 bar in an NPT ensemble using a Nosé-Hoover thermostat [Evans and Holian, 1985, Nosé, 1984] and a Parrinello-Rahman barostat [Parrinello and Rahman, 1981]. For each system, a 1 μs MD simulation was performed with Gromacs 2018 [Abraham et al., 2015]. For the mixed-solvent MD simulations, benzene molecules were used as probes with a benzene concentration of 1 M. Benzene was parameterized with GAFF. To retain diffusive behavior and to prevent phase separation due to hydrophobic interactions, an interligand repulsion term was added manually to the non-bonded parameters of the probe molecules with $\sigma = 8$ Å and $\varepsilon = 1.5{\cdot}10^{-3}\frac{kJ}{mol}$ for carbon atoms. For each system, 20 replicates of mixed-solvent MD simulations with 500 ns each were performed with Gromacs 2016.

6.2.2 SWISH simulations

Equilibrated protein structures served as starting structures for SWISH simulations. SWISH simulations, with 6 replicas each, were performed with the HREX implementation of the Plumed 2.4 plugin for Gromacs 4.6.7. The different non-bonded interactions between protein and water molecules were edited manually in the topology. Non-bonded parameters between the oxygen atoms of water molecules and carbon/sulfur atoms of the protein were calculated for each replica according to the Lorentz-Berthot mixing rules and scaled from 1 to 1.35. An exchange between two replicas was attempted every 2 ps. The total simulation length for each system was 500 ns.

6.2.3 Analysis

The pocket volume was calculated with the VMD plugin Epock [Laurent et al., 2015]. The radius and shape was chosen based on visual inspection after alignment of apo and ligand-bound structures (Table 6.2, Fig. 6.1). The van der Waals radius for the grid sphere size to determine the free space, was set to 1.4 Å with a cutoff of 4 Å and a grid resolution

protein	shape	target 1	target 2	radius [Å]	vol. apo [Å³]	vol. holo [Å³]
NPC2	cylinder	Val20 Val38	Val64 Tyr100	4.0	30.38	429.25
p38α	sphere	Ile193 Lys249		5.4	58.00	231.88
		Ser251 Ile259				
		Leu291				
ricin	sphere	Leu74 Asn78		3.0	0.00	64.00
		Asn122 Tyr123				
		Ile172 Asn209				
Eg5	sphere	Arg119 Ile136		5.0	54.62	221.75
		Glu215				

Table 6.2: Pocket definitions and volume (vol.) for apo and ligand-bound structures for NPC2, p38α, ricin, and Eg5.
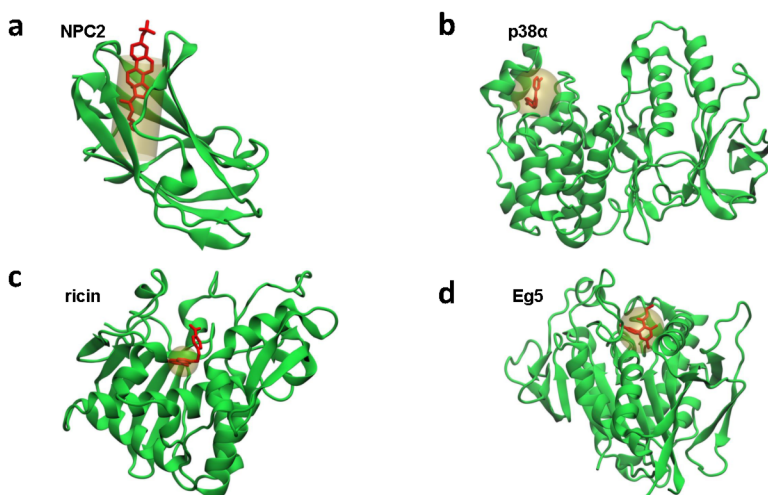
Figure 6.1: Pocket definitions (transparent yellow) of each protein (green) enclosing the respective ligands (red). Taken from Tiede [2021].

of 1 Å.

6.3 Results and discussion

6.3.1 Pocket opening mechanisms in NPC2, p38α, ricin, and Eg5

In all systems, NPC2, p38α, ricin, and Eg5, pocket openings could be observed. At least one method, of all three simulation-based approaches, was capable of sampling the cryptic pocket opening fully and over a considerable duration without a ligand. The characteristics of the different simulation approaches are important to further understand the binding mechanism. Depending on which method performs best, the respective binding mechanism can be placed in the spectrum of induced-fit and conformational selection.

In the case of NPC2, mixed-solvent MD simulation with 1 M benzene molecules was the best suited method, because it sampled pocket openings efficiently. The highest mean value for pocket opening was observed in replicas 7, 8, 11, and 14 (Fig. 6.2). These replicas showed the binding of two benzene molecules to the pocket, indicating that the binding pocket can accommodate a ligand with two aromatic rings. The benzene molecules insert sequentially into the hydrophobic pocket. A benzene molecule binds first into a surface-exposed binding
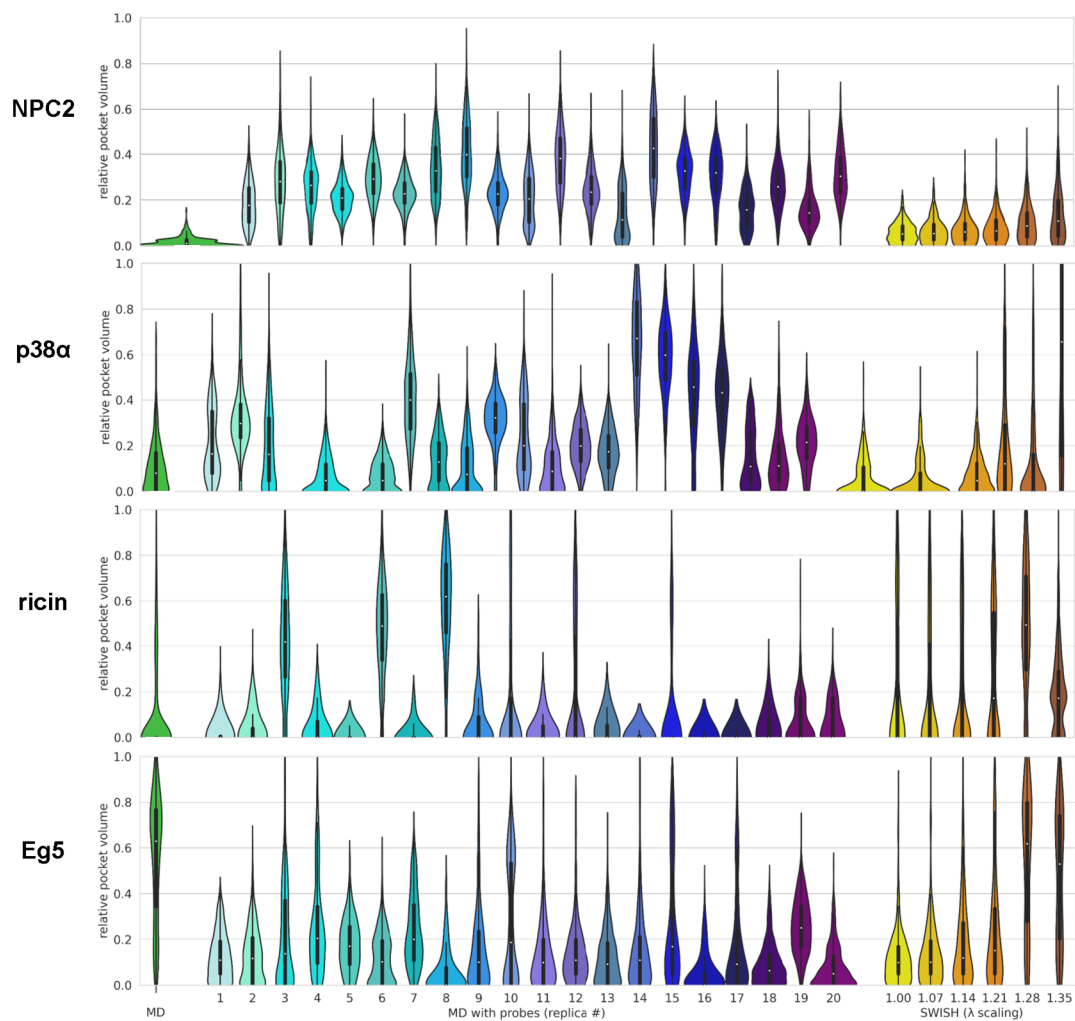
Figure 6.2: Violin plots show the distribution of the relative pocket volume in reference to the ligand-bound structure for MD simulations with and without probes and the SWISH simulation of NPC2, p38α, ricin, and Eg5. Adapted from Tiede [2021].
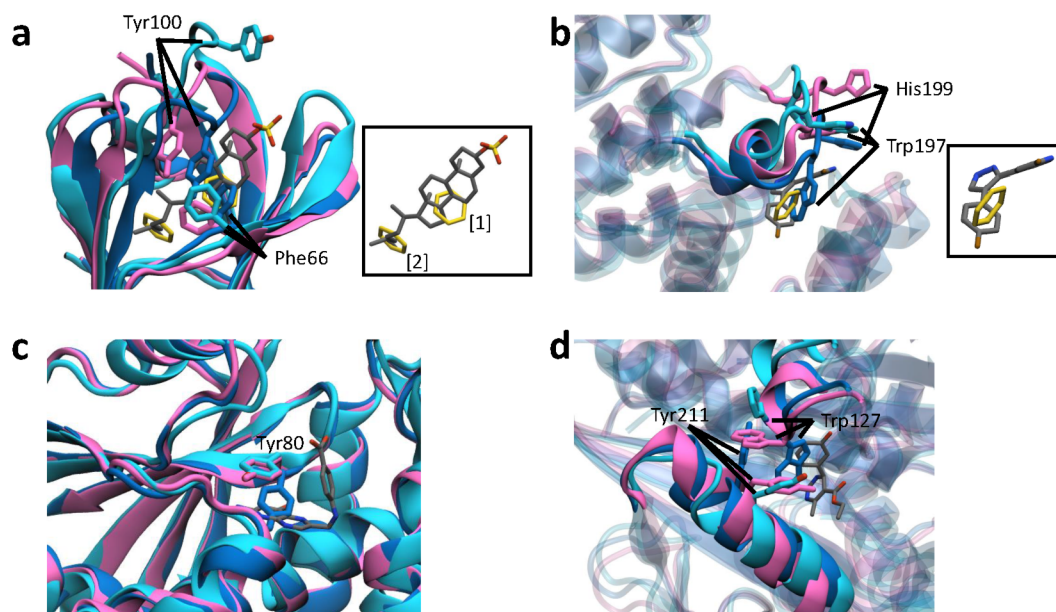
Figure 6.3: Superposition of apo X-ray structure (blue), holo X-ray structure (magenta) and snapshot from MD simulation (cyan). Pocket residues and ligand (grey) are shown with atomic detail. **a,** NPC2: snapshot from mixed-solvent MD simulation (replica 14) at 470 ns. **b,** p38α: snapshot from mixed-solvent MD simulation (replica 14) at 480 ns. **c,** ricin: snapshot from brute-force MD at 1 μs. **d,** Eg5: snapshot from brute-force MD at 950 ns.

site ([1] for NPC2 in Fig. 6.4a) between the two beta strands and then moves deeper into the pocket ([2] for NPC2 in Fig. 6.4a), making room for a second benzene molecule. The two benzene molecules overlay well with the known cryptic pocket binder and stabilize the pocket residues in their open conformation (Fig. 6.3a). In all 20 replicas, 25% of the simulation time two benzene molecules were bound rigidly, 65% of the time one benzene molecule was bound and only 10% of the time the pocket was unoccupied. Both in the brute-force MD and SWISH simulation, complete pocket opening was not observed. During brute-force MD simulation, the cryptic pocket remained closed. The Tyr100 side chain, located on the protein surface, showed some flexibility, but Phe66 retained its starting conformation and kept the pocket occluded. Similar to the observations from the brute-force MD run, the cryptic pocket did not open up fully in the SWISH simulation. Side chain flipping of Tyr100 was observed, with increased frequency with increasing bias, while the other residues in the pocket remained mostly in their starting conformation. Hence, only Tyr100 was affected by the scaled water-protein interaction, while the scaling had no effect on the more buried residues, including Phe66. Because the pocket only fully opened upon probe binding, induced-fit very likely plays a major role in the binding mechanism. Further, by looking at all available unbound X-ray crystal structures of NPC2, Beglov et al. [2018] found no evidence that the site can open without ligand binding, in contrast to several proteins in the CryptoSite set. Comitani and Gervasio [2018] also investigated NPC2 with the same simulation approaches as used in this Thesis. They observed similar results for the brute-force MD and SWISH simulation, however, in contrast to our observations they did not obtain full pocket opening with MD simulations with probe molecules. Nevertheless,

they also observed two benzene binding positions in the cryptic site, in agreement with our results.

For p38α, mixed-solvent MD simulations led to efficient pocket openings and showed that one benzene molecule can bind to the pocket. The benzene molecule can enter the pocket from multiple directions and, when bound, is not fixed at one specific position, but is able to move slightly in the pocket. The bound benzene molecule affects the conformation of the key residue Trp197. Depending on its location in the pocket, it pushes the Trp197 side chain to a partially or fully open conformation. For 92% of the simulation time, the pocket was accommodated by one benzene molecule. Unlike NPC2, the brute-force MD simulation also showed short partial openings and some flexibility of the binding site residues. In the first 20 ns, the Trp197 loop moved outwards, resulting in an almost opened pocket. However, in the remaining trajectory Trp197 did not completely obtain its holo conformation and the His199 loop was only flexible in few time frames, displaying a rather rigid pocket. Moderate flexibility of Trp197 in the brute-force MD simulation and efficient pocket openings by benzene binding suggests a combination of induced-fit and conformational selection in the ligand binding mechanism. The SWISH simulation did only sample occasional pocket openings in replicas 3 and 4 and the protein structure completely unfolded in replica 5. The replica exchange between replica 4 and 5 stopped at 110 ns, indicating the biasing parameters were not suitable for p38α. Our observations agree with the results from Comitani and Gervasio [2018], except for the SWISH simulation, where slightly less pocket openings were observed, explained by with the stop in replica exchange.

Regarding ricin, the SWISH simulation performed best and led to full pocket openings in all replicas. The side chain orientation of the pocket residue Tyr80 primarily determines if the pocket is occluded or open. The SWISH simulation sampled various Tyr80 conformations and with increasing bias Tyr80 spent more simulation time in its open conformation. In replica 5, some protein unfolding was observed, but this did not affect the binding site. In the brute-force MD run, the pocket also opened up. At 888 ns, Tyr80 flipped from its closed to its open conformation and remained in the flipped-out conformation, resulting in pocket opening. Unlike NPC2 and p38α, benzene molecules did not bind to the cryptic pocket and had no effect on pocket formation. Only in one replica of the mixed-solvent MD simulations, in replica 8, a benzene molecule temporarily interacts with the pocket residue Tyr80, leading to side chain flipping and pocket opening, but the molecule does not overlay with the known binder. Most benzene molecules bound not directly in the cryptic pocket, but more towards the surface and did not displace the Tyr80 side chain. In several replicas, no benzene molecules bound to or close to the cryptic pocket, and nevertheless pocket openings were observed. Because pocket openings were observed independent of benzene binding, a major role of conformational selection in the ligand binding mechanism is proposed.

In the case of Eg5, the pocket was open most of the time in the brute-force MD simulation. Overall, conformational changes of Trp127 side chain contribute most to the pocket opening. In the brute-force MD run, all pocket residues displayed some flexibility. At 190 ns, Trp127 flipped outwards, resulting in bigger pocket volume for the remaining trajectory.

In addition, Tyr211 transitioned to its open conformation at 870 ns, leading to complete pocket opening. Glu116 already obtained its open conformation after energy minimization and retained it for the whole simulation time. The mixed-solvent MD simulation found a probe binding site within the cryptic cavity, but it was only occupied rarely, compared with NPC2 and p38α. Benzene molecules more often entered into another binding pocket beneath the known cryptic pocket (see Chapter 6.3.2), not resulting in pocket opening. The cryptic pocket of interest was occupied for only 16% of the simulation time. Pocket openings were also observed in the SWISH simulation and occurred more frequent and longer with the increase in bias. Because pocket opening was sampled efficiently with brute-force MD and benzene molecules binding played only a minor role, conformational selection probably dominates the ligand binding mechanism.

6.3.2 Mixed-solvent MD simulations reveal potential unknown binding pockets

Mixed-solvent MD simulations show various binding pockets for benzene molecules in all four systems (Fig. 6.4). In NPC2 and p38α, only known binding pockets accomodate benzene molecules. For NPC2, both binding positions for probe molecules are located in the cryptic pocket. For p38α, the cryptic pocket and the ATP pocket can be populated with probe molecules. New binding pockets were found for ricin and Eg5. In ricin, four benzene binding pockets were observed. The most occupied position is close to the known cryptic pocket [2], but not directly where the known ligand sits [1]. Two additional binding positions [3,4] were often occupied in multiple replicas, with a residence time of at least 100 ns. One benzene binding spot is between Phe119 and Phe168 and close to Phe93 [3] and the other one is between Leu207 and Phe240 [4]. In Eg5, benzene molecules bound frequently to a binding pocket beneath the known cryptic site [2] close to Leu171 and Leu214. In addition, benzene molecules were stably bound in multiple replicas for at least 100 ns at sites distant from the cryptic pocket, close to Phe144 and Val210 [3], between Ile196, Ile319 and Val238 [4] and between Val41 and Val71 [5].

6.4 Summary and conclusion

Here, I used three simulation-based approaches to study the dynamical properties of cryptic pockets in four protein drug targets: NPC2, p38α, ricin, and Eg5. For all proteins, X-ray crystal structures of the unbound and ligand-bound conformations are available. Starting from the unbound structure, I used MD, mixed-solvent MD, and SWISH simulations to investigate the opening mechanisms of the binding pockets.

We were able to resolve the opening of the cryptic pockets for all four proteins. Different simulation methods enabled us to gain insights into their respective opening mechanisms and put them in the spectrum of induced-fit and conformational selection. In the mixed-solvent MD simulations of NPC2, two benzene molecules consecutively inserted into the cryptic site. The pocket opening was correlated with benzene binding, which indicates that an induced-fit mechanism plays a dominant role in the ligand binding process. The cryptic pocket of p38α, on the other hand, showed flexibility even in the absence of probe molecules.
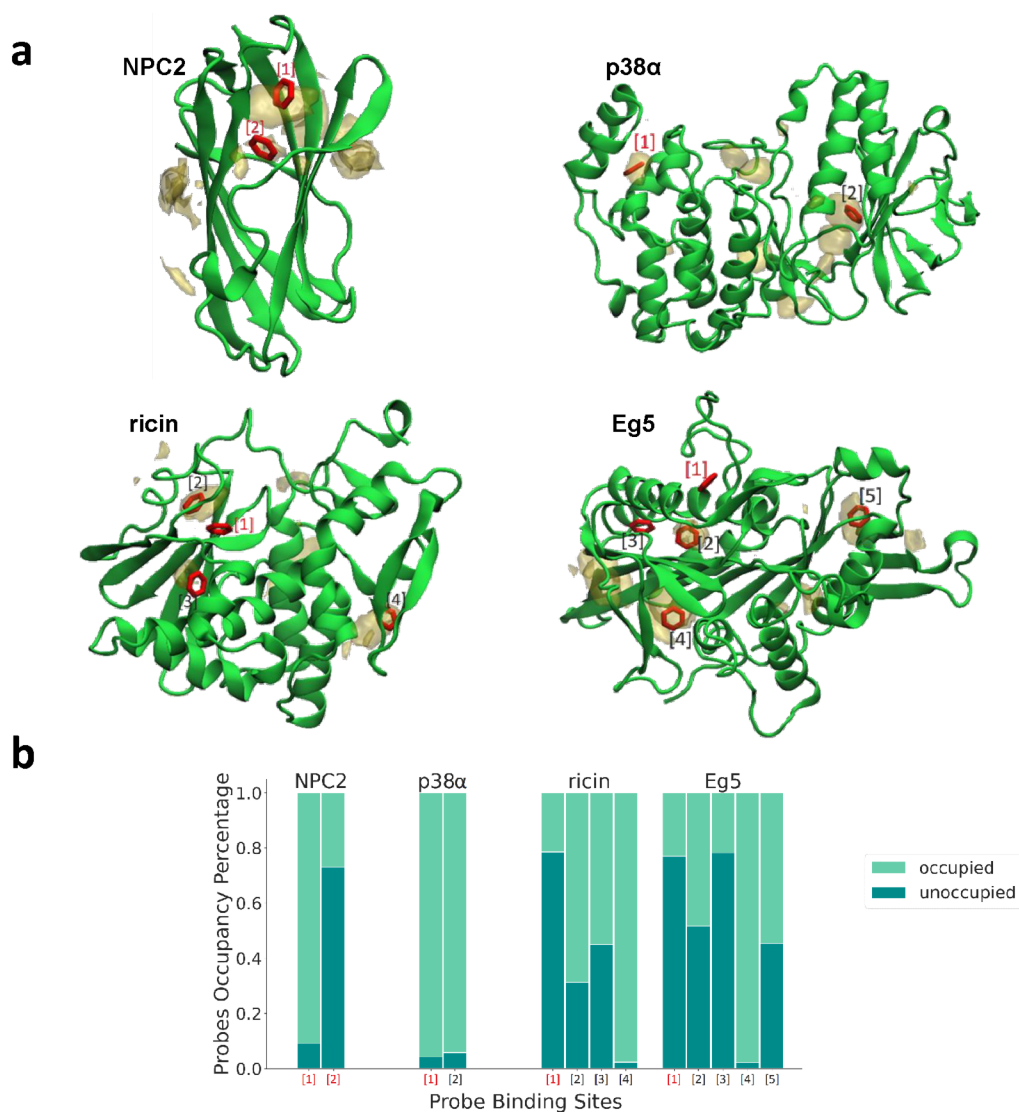
Figure 6.4: **Additional probe binding sites. a,** Probe densities (yellow), indicated with brackets [], found with mixed-solvent MD simulations for NPC2, p38α, ricin and Eg5. **b,** Probes occupancy percentage for each binding site of NPC2, p38α, ricin, and Eg5. Benzene binding spots in in the cryptic pocket under consideration (red numbers), and in additional protein pockets (black numbers). Taken from Tiede [2022].

Also, in the mixed-solvent MD simulations of p38α, one benzene molecule frequently entered into the cryptic pocket, suggesting a joint effect of induced-fit and conformational selection. While the conformation of cryptic pockets in NPC2 and p38α was greatly affected by probe molecules, those of ricin and Eg5 were not. Hence, I propose a major role of conformational selection in the ligand binding mechanism of ricin and Eg5. In addition to the previously known cryptic binding sites, several new binding pockets were observed in ricin and Eg5 by benzene binding.

In general, we observed that the water molecules with scaled interactions of the SWISH simulations can enhance the conformational sampling of solvent-exposed residues, while they seem not to be suited to enter into deep cavities, e.g., in NPC2. In this case, mixed-solvent MD simulations seem to perform best. Comitani and Gervasio [2018] also combined SWISH with probe molecules, which improved the sampling in NPC2. As seen in the SWISH simulation of p38α, the scaling factor λ needs to be chosen carefully, because the protein unfolded in the most biased replica. This makes the wide application of SWISH difficult. Further, our simulations show that the mechanism for cryptic pocket opening is system dependent.

Chapter 7

Preferential substrate interactions of papain-like protease in SARS-CoV-2

7.1 Introduction

The coronavirus disease 19 (COVID-19) escalated into a global pandemic in 2020 caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) [Huang et al., 2020]. Coronaviruses belong to a large family of enveloped single-stranded RNA viruses and can cause respiratory, hepatic, gastrointestinal, and neurologic diseases in both animals and humans [Weiss and Leibowitz, 2011]. To date, there have been seven human coronaviruses identified, including Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) [Drosten et al., 2003], and Middle East Respiratory Syndrome Coronavirus (MERS-CoV) [Zaki et al., 2012]. Due to the vast distribution of coronaviruses, the huge genetic variation, frequent recombination of their genomes, and the increase of human-animal interface activities, new coronaviruses emerge sporadically in humans [Cui et al., 2019, Zhu et al., 2020].

In contrast to the related SARS-CoV, responsible for the SARS outbreak in 2003, the symptoms of COVID-19 are mostly less severe and the case-fatality rate is lower. However, SARS-CoV-2 has a much higher transmission rate [Wu et al., 2020, Yuen et al., 2020]. Both viruses are vitally dependent on the activity of two proteases, namely the main protease and the papain-like protease (PLpro). These proteases process the polyproteins that are translated from the viral RNA. In addition to the polypeptide processing, the PLpro modulates the host immune system via deubiquitination and deISGylation of host cell proteins, resulting in inhibition of the host antiviral innate immune response [Frieman et al., 2009]. As a result, PLpro is an important therapeutic target to halt virus replication.

Recently, we found with our collaborators that inhibiting PLpro not only stops viral replication, but also improves the antiviral immune response [Shin et al., 2020]. Upon infection, SARS-CoV-2 must overcome numerous human defense mechanisms, including the innate immune defense. The innate immune system functions as the first line of host defense against pathogens, and limits viral invasion or replication. Natural killer cells can sense messenger molecules called type I interferons released by infected body cells, which leads to depletion of infected cells. By cleaving proteins of infected cells, PLpro can disturb the release of type I interferons and in effect can block the innate immune response of the host (Fig. 7.1).

Shin et al. [2020] also discovered that the viral protein PLpro of SARS-CoV-2 (SCoV2-PLpro) cleaves ISG-15 (interferon-stimulated gene-15) from cellular proteins with greater activity than the SARS equivalent (SCoV-PLpro), leading to reduced interferon response. This
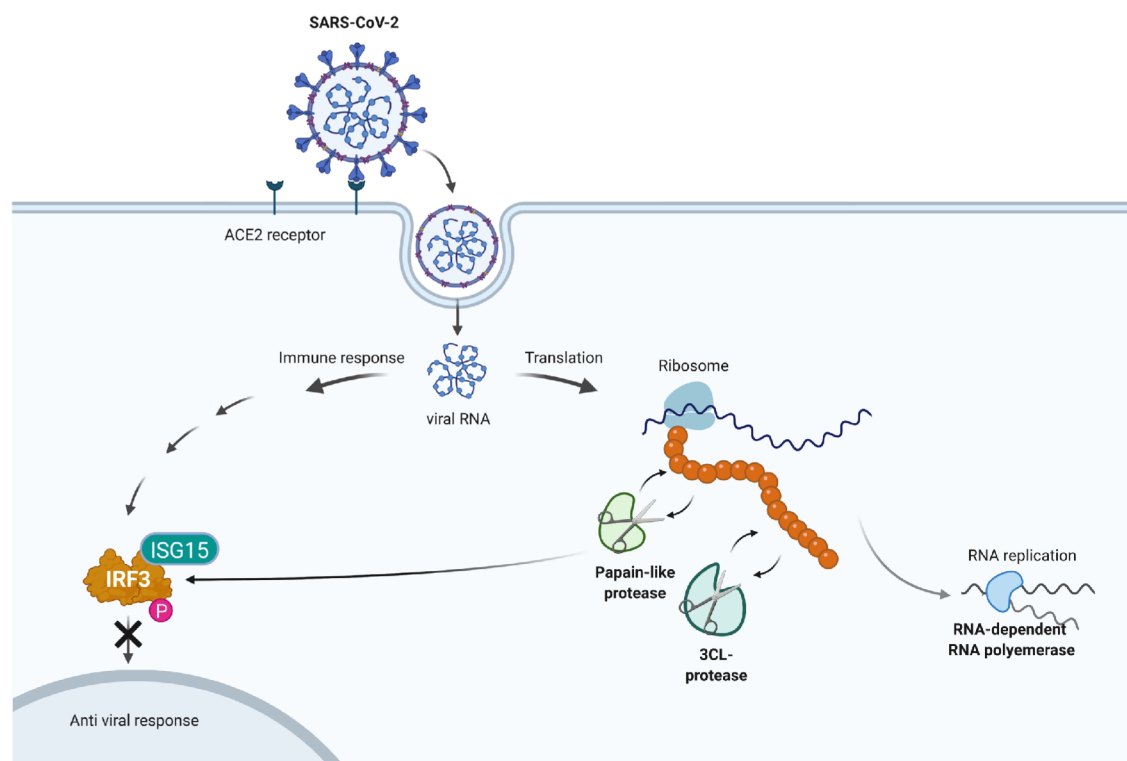
Figure 7.1: **Schematic representation of the role of SCoV2-PLpro in the viral life cycle.** The role of SCoV2-PLpro in both polypeptide processing and host-immune response is depicted. Created with BioRender.com.

makes SARS-CoV-2 more contagious than the SARS equivalent, even though SCoV-PLpro and SCoV2-PLpro share an 82.9% sequence identity. To understand the underlying mechanism for the difference in substrate interaction, I performed MD simulations of PLpro of SARS-CoV and SARS-CoV-2 with both mISG15 and ubiquitin (Ub).

Due to the urgent need of therapies for the pandemic, repurposing approved drugs is a promising strategy. The small molecule GRL-0617 is a potent, selective and competitive noncovalent inhibitor of SCoV-PLpro. It was found in several studies that GRL-0617 can also inhibit SCoV2-PLpro [Shin et al., 2020, Freitas et al., 2020, Klemm et al., 2020, Fu et al., 2021]. With MD simulations, I was able to predict and investigate its binding mode.

7.2 Methods

7.2.1 MD simulations

In total, seven different MD systems were set up. Both SCoV-PLpro (WT and double mutant Ser67Val/Leu76Thr) and SCoV2-PLpro with bound K48-Ub2 systems were set up. Starting from the coordinates of PDB entry: 5e6j [Békés et al., 2016], the mutations Ser67Val and Leu76Thr were introduced using Modeller [Šali and Blundell, 1993]. For SCoV2-PLpro with bound K48-Ub2, two models were built. I combined the substrate coordinates taken from PDB entry: 5e6j with (1) the X-ray crystal structure of the apo form (PDB entry: 6w9c [Osipiuk et al., 2020], re-refined by T. Croll (University of Cambridge, UK) (https://drive.google.com/drive/folders/1JBo50CdkBU7K1pFThuqrzhQ-NcsIAWyG)

and (2) with the X-ray crystal structure of the SCoV2-PLpro–mouse ISG15 complex (PDB entry: 6yva [Shin et al., 2020]) after PLpro alignment using PyMol. In all di-ubiquitin systems, the triazole linker was substituted with Lys using Modeller and a harmonic-distance restraint potential between the backbone carbonyl carbon atoms of Lys48 and Gly75 with a target distance of 9.5 Å and a force constant of 502.080 $\frac{kJ}{mol \cdot nm^2}$. The covalent propargylamide linker was removed. SCoV2-PLpro with bound mISG15 was built based on the X-ray crystal structure of the SCoV2-PLpro–mouse ISG15 complex (PDB entry: 6yva). Missing residues of SCoV2-PLpro and one Zn^{2+} ion were modelled according to the X-ray crystal structure (PDB entry: 6w9c, re-refined by T. Croll). In addition, two inhibitor-bound complexes were set up, the inhibitor GRL-0617 is once bound to SCoV-PLpro and once to SCoV2-PLpro. The coordinates of the SCoV-PLpro–GRL-0617 complex were taken from PDB entry: 3e9s [Ratia et al., 2008]. The oxidized Cys112 was switched to the reduced form (SH) using Modeller. The simulation model of the SCoV2-PLpro–GRL-0617 complex was built according to the X-ray structure of the apo form of SCoV2-PLpro (PDB entry: 6w9c, re-refined by T. Croll). The compound GRL-0617 was manually placed into the binding site according to PDB entry: 3e9s after PLpro alignment using PyMol. The blocking loop 2 (BL2 loop, GNYQCGH) capping the GRL-0617 binding site was remodelled according to the SCoV-PLpro X-ray crystal structure of the complex (PDB entry: 3e9s) using Modeller. The GRL-0617 ligand was parameterized with GAFF [Wang et al., 2004].

In all set-ups, I set up all ionizable residues in their physiological protonation state, except His17 of SCoV2-PLpro (His18 of SCoV-PLpro) and His272 of SCoV2-PLpro (His273 of SCoV-PLpro) were charged. Missing side chains were modelled using Modeller and all crystallographic water molecules and ions were retained, except a nickel ion in PDB entry: 5e6j. I used Gromacs 2018 [Abraham et al., 2015] for the MD simulations with the AMBER99SB*-ILDN-q force field [Hornak et al., 2006a, Best and Hummer, 2009, Best et al., 2012, Lindorff-Larsen et al., 2010] combined with TIP4P-D water [Piana et al., 2015] and 150 mM NaCl. After energy minimization, I performed five rounds of equilibration, in which I gradually weakened the position restraints on protein heavy atoms, first in an NVT ensemble (0.25 ns) and then in an NPT ensemble (4 x 0.5 ns) using a Berendsen thermostat and barostat [Berendsen et al., 1984]. Production simulations were run at 310 K and at a pressure of 1 bar in an NPT ensemble using the Nosé–Hoover thermostat [Evans and Holian, 1985, Nosé, 1984] and the Parrinello-Rahman barostat [Parrinello and Rahman, 1981]. I set up three independent runs of the SCoV2-PLpro systems with bound substrates, starting from different Modeller results for the apo-like model of SCoV2-PLpro:K48-Ub2 and for the SCoV2-PLpro–mouse ISG15 complex.

7.2.2 Analysis

For simulations with bound substrates and with bound inhibitor, I calculated the RMSD of each backbone substrate (distal ubiquitin in K48-Ub2 and N-terminal domain of mouse ISG15) and of GRL-0617 (heavy atoms) with respect to the equilibrated structure after alignment on the helix backbone of PLpro (without the flexible UBL domain). From simulations of SCoV-PLpro–K48-Ub2, the minimum heavy-atom distance between Phe70
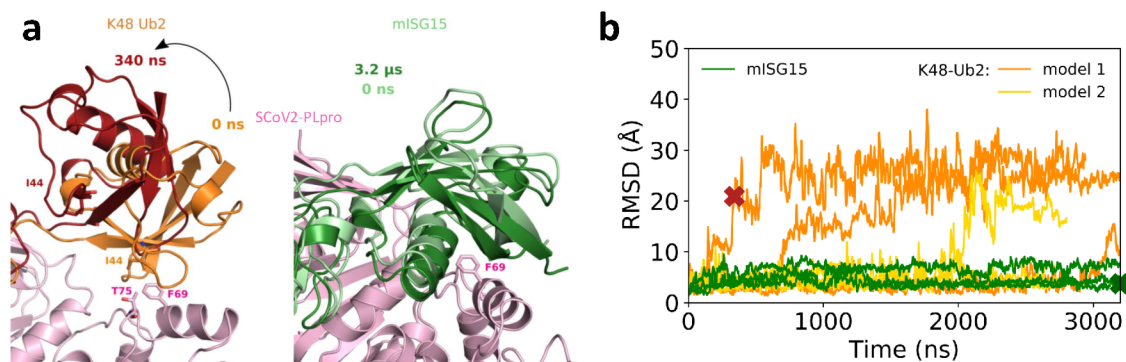
Figure 7.2: **SCoV2-PLpro interacts more tightly with ISG15 compared with K48-Ub2.** Adapted from Shin et al. [2020], with kind permission by Springer Nature. **a,** Snapshots from MD simulations of SCoV2-PLpro (light pink cartoon) with (left) K48-Ub2 at 340 ns and (right) mISG15 at 3.2 μs. Key residues in the interface are highlighted. **b,** Backbone RMSD of the N-terminal domain of mISG15 (green) and of the distal ubiquitin in K48-Ub2 in an apo-like model (orange, model 1, SCoV2-PLpro coordinates from substrate unbound form, PDB entry: 6w9c) and in an mISG15-like model (yellow, model 2, SCoV2-PLpro coordinates from substrate bound form, PDB entry: 6yva) from their respective SCoV2-PLpro-bound starting structures as function of time. The RMSD was calculated after superimposing the helix backbone atoms of SCoV2-PLpro. Time points for structural snapshots are marked with a cross.

of PLpro and Ile44 of ubiquitin was monitored.

7.3 Results and discussion

7.3.1 SCoV2-PLpro interacts more tightly with ISG15 compared with K48-Ub2

Experimental studies showed that SCoV-PLpro and SCoV2-PLpro exhibit differences in their substrate preferences [Shin et al., 2020]. SCoV2-PLpro preferentially cleaves ISG15 over ubiquitin chains, whereas SCoV-PLpro strongly reduces ubiquitination. To find an explanation for the different substrate preferences, I set up MD simulations of SCoV-PLpro and SCoV2-PLpro with K48-Ub2 and mouse ISG15. Simulations of SCoV2-PLpro clearly showed that SCoV2-PLpro interacts more tightly with ISG15 compared with K48-Ub2, which is consistent with the experimental findings. While ISG15 stayed bound in three independent runs of 3.2 μs, as observed in the SCoV2-PLpro–ISG15 X-ray structure, the distal ubiquitin of K48-Ub2 dissociated from SCoV2-PLpro in four out of six runs (Fig. 7.2). Because no X-ray crystal structure of the SCoV2-PLpro with K48-Ub2 was available, I built two models by combining the ubiquitin coordinates from the SCoV-PLpro ubiquitin complex X-ray crystal structure (PDB entry: 5e6j) with the SCoV2-PLpro coordinates after PLpro alignment. The PLpro coordinates were once taken from the X-ray crystal structure of the unbound form (PDB entry: 6w9c, model 1) and once from the X-ray crystal structure of the ISG15-bound complex (PDB entry: 6yva, model 2). With both models as initial structures, I could observe ubiquitin separation. For model 1, in all three runs ubiquitin separated from SCoV2-PLpro and for model 2 one out of three runs showed ubiquitin dissociation. Taken together, I observed a tighter interaction between SCoV2-PLpro with ISG15 compared with
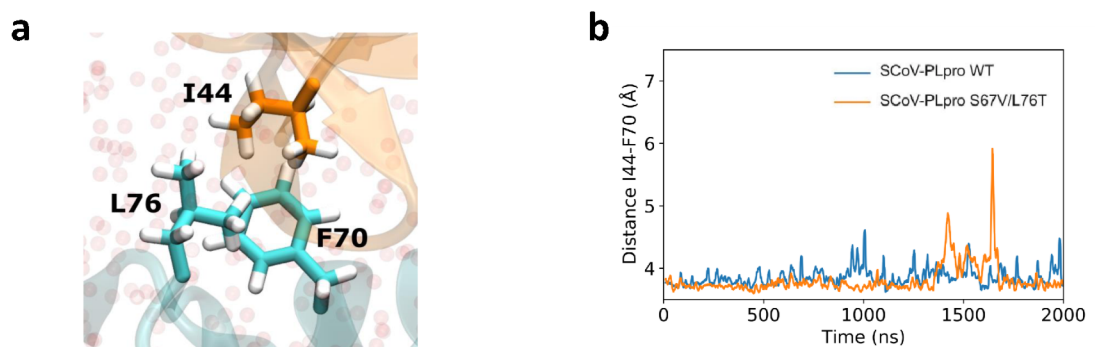
Figure 7.3: **SCoV-PLpro interacts more tightly with ubiquitin compared with SCoV2-PLpro.** Adapted from Shin et al. [2020], with kind permission by Springer Nature. **a,** X-ray crystal structure of SCoV-PLpro:K48-Ub2, focusing on the hydrophobic triad Phe70, Leu76(PLro) and Ile44(Ub). Red spheres denote water oxygens. **b,** Minimum heavy atom distance between Phe70(PLro) and Ile44(Ub) in WT and double mutant (Ser67Val/Leu76Thr) of SCoV-PLpro:K48-Ub2 as function of time.

ubiquitin. PLpro enzymes from both previous and current SARS viruses have the same core residue that mediates hydrophobic interactions with either ubiquitin or ISG15: Phe69 in SCoV2-PLpro and Phe70 in SCoV-PLpro. In contrast to ubiquitin, ISG15 has a second hydrophobic binding interaction with SCoV2-PLpro at Val66 (mutated in SCoV-PLpro to Ser67), which might be a reason for the tighter ISG15 interaction in SCoV2.

To understand the difference in ubiquitin substrate preference between PLpro from the previous and current SARS virus, I performed an MD simulation of the SCoV-PLpro bound to ubiquitin. In contrast to the SCoV2-PLpro simulations, PLpro from the previous SARS virus showed a tight interaction to the ubiquitin substrate. Zooming into the binding interface, I observed a tight hydrophobic cluster consisting of Leu76, Phe70 (PLpro) and Ile44 (Ub) (Fig. 7.3a). In a second simulation, I mutated two residues in SCoV-PLpro *in silico* to resemble the binding interface of SCoV2-PLpro. Indeed, I observed that the SCoV-PLpro double mutant (Ser67Val/Leu76Thr) with K48-Ub weakened the hydrophobic cluster within the binding interface and showed transient ubiquitin dissociation (Fig. 7.3b). Because Leu76 is part of the crucial hydrophobic cluster within the binding interface and Ser67 has no interaction with ubiquitin, we identified Leu76Thr as the key mutation between PLpro from the previous and current SARS virus, that changes the ubiquitin cleavage activity.

7.3.2 Dissociation mechanism of ubiquitin

All dissociation events of ubiquitin from SCoV2-PLpro show the same pattern. Visual analysis of the atomistic trajectories revealed that separation is initiated by water molecules that transiently enter between K48-Ub2 Ile44 and SCoV2-PLpro Thr75. Water molecules probed this interaction, until water molecules wedge in between K48-Ub2 Ile44 and SCoV2-PLpro Thr75. This weakens the hydrophobic interaction and water molecues can penetrate between Thr75/Phe69(SCoV2) and Ile44(Ub), which results in dissociation. The MD simulations show that the dissociation pathway is water mediated (Fig. 7.4).
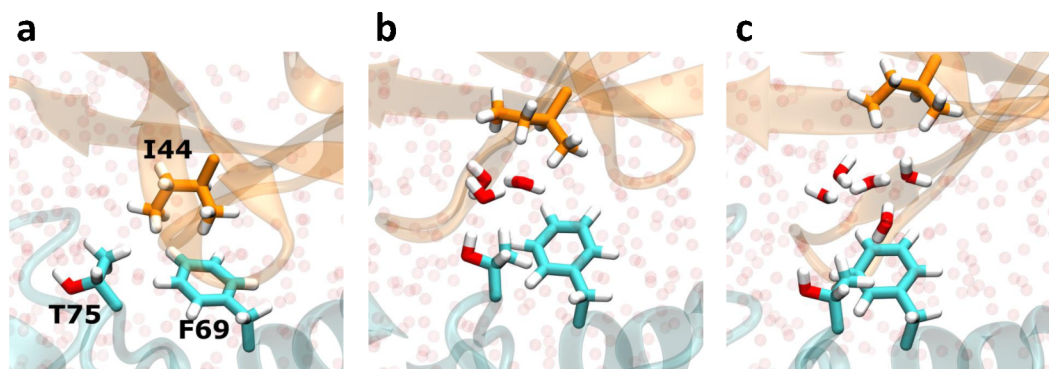
Figure 7.4: **Water mediated dissociation pathway.** Adapted from Shin et al. [2020], with kind permission by Springer Nature. **a,** Initial hydrophobic interactions between Phe69(SCoV2), Thr75(SCoV2) and Ile44(Ub). **b,** Water wedges in between Thr75(SCoV2) and Ile44(Ub). **c,** Water penetration between Thr75/F69(SCoV2) and Ile44(Ub) leads to dissociation.

7.3.3 Binding mode of GRL-0617 is the same between SCoV-PLpro and SCoV2-PLpro

At the time this research collaboration was conducted, no structure of SCoV2-PLpro in complex with the compound GRL-0617 was available. To understand the mechanism of inhibition of SARS-CoV-2 by GRL-0617, I modelled the compound into the apo form of SCoV2-PLpro, using as template the X-ray co-crystal structure of GRL-0617 bound to SCoV-PLpro (PDB entry: 3e9s), and changed the blocking loop 2 (BL2) capping the GRL-0617 binding site according to the bound form of SCoV-PLpro X-ray crystal structure (PDB entry: 3e9s). The binding site shares 100% sequence identity with the SARS equivalent [Ibrahim et al., 2020]. The MD simulations revealed that the complexes are stable and that the ligands did not deviate significantly from the starting structures, having RMSD values smaller than 2 Å (Fig. 7.5a,b). Because MD simulations of both complexes showed that they are equally stable during the run time of 1 µs, the modelled binding mode of GRL-0617 to SCoV2-PLpro resembles the actual binding mode. Several recent studies confirmed this finding by solving the X-ray co-crystal structure of SCoV2-PLpro in complex with the compound GRL-0617 [Osipiuk et al., 2021, Gao et al., 2021, Fu et al., 2021, Ma et al., 2021]. The X-ray crystal structures overlay well with the predicted conformation from simulation (Fig. 7.5c). Taken together, MD simulations and the newly solved X-ray crystal structures observed a common binding mode of GRL-0617 to both SCoV-PLpro and SCoV2-PLpro.

The inibitor GRL-0617 is stabilized by the capped BL2 loop (Fig 5 d). Two hydrogen bonds are formed between the BL2 loop and the ligand. The sidechain oxygen of Tyr268 interacts with the amino group on the benzene ring of GRL-0617, and the backbone amino group of Gln269 forms a hydrogen bond with the carbonyl oxygen of GRL-0617. In addition, several π-π stacking interactions between the Tyr268 and the naphthalene and benzene group of GRL-0617 rigidify the compound (Fig. 7.5d).
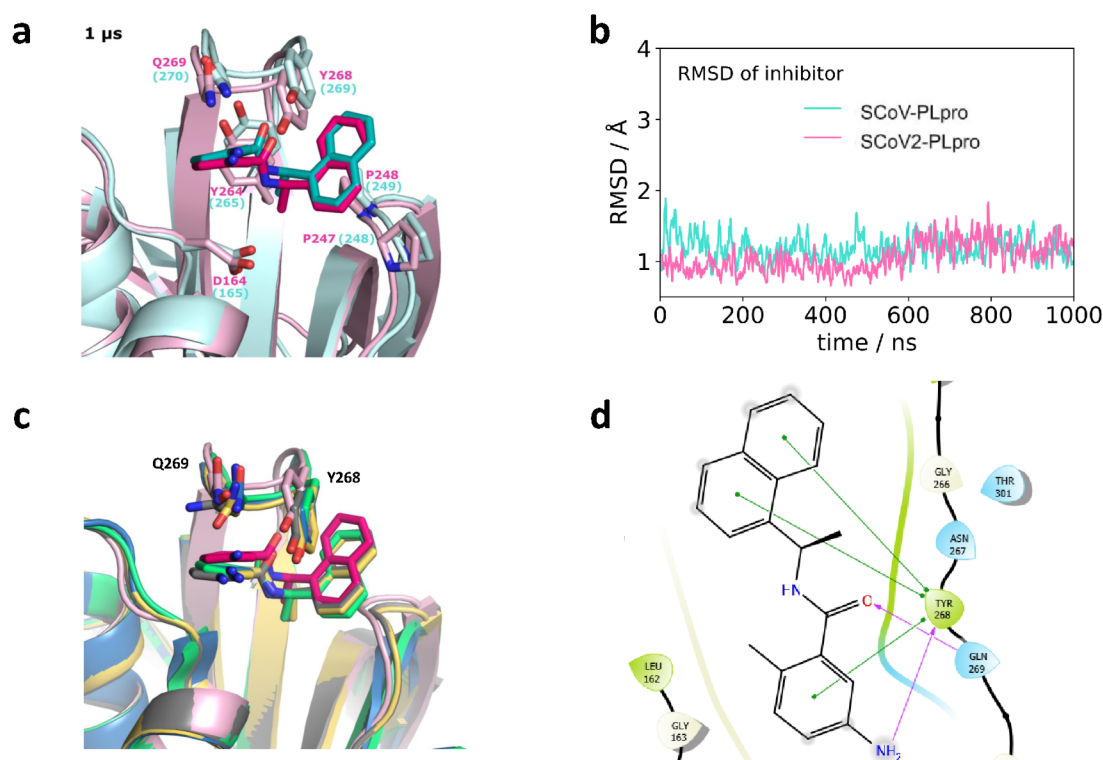
Figure 7.5: **Binding pocket of SCoV2-PLpro is preserved.** a, and b, reprinted from Shin et al. [2020], with kind permission by Springer Nature. **a,** Snapshots of SCoV-PLpro (light cyan) and SCoV2-PLpro (light pink) with bound GRL-0617 (dark colours) after 1 µs of MD simulation. The protein backbones are shown in cartoon representation, and the ligand with contacting residues as sticks. **b,** RMSD of the GRL-0617 bound to SCoV-PLpro (light blue) and SCoV2-PLpro (light pink) as a function of time. The RMSD was calculated for non-hydrogen atoms of GRL-0617 with respect to the starting structures in the MD simulations after superimposing the helix backbone atoms of PLpro. **c,** Superposition of 1 µs MD frame of SCoV2-PLpro (light pink) with bound GRL-0617 (dark pink) with X-ray crystal structures of SCoV2-PLpro GRL-0617 complexes: PDB entries: 7jrn (gray), 7cmd (yellow), 7jir (blue), 7cjm (green). **d,** Ligand interaction diagram for GRL-0617 from PDB entry: 7cjm.

7.4 Summary and conclusion

With MD simulations, I found that the viral protein PLpro of SARS-CoV-2 interacts more strongly with ISG15 compared to ubiquitin, whereas PLpro from the SARS equivalent tightly interacts via a hydrophobic cluster with ubiquitin. This can explain the differences in substrate activities observed in experiments. The MD trajectories gave us a unique opportunity to explore the origin of the differences in specificity. Inspection of simulation data revealed that the disruption of a hydrophobic triad in the core of the K48-Ub2 binding interface is presumably the main factor, primarily due to the mutation of Leu76(SCoV) to Thr75(SCoV2). The model of SCoV2-PLpro bound with ubiquitin based on the apo form potentially underestimates the interaction, because full dissociation occured in all three runs, whereas the model based on the substrate-bound form showed one full dissociation of the distal ubiquitin and two transient partial separations to an RMSD of 10 Å. Nevertheless, the subtle change in the binding interface clearly reduces the interaction between PLpro and ubiquitin. All dissociations were initiated by water penetrating between Ile44(Ub) and Thr75(SCoV2). Therefore, consistently and independent of the initial model, the observed dissociations occur with the same mechanism of water first entering into the hydrophobic contact between Ile44(Ub) and Thr75(SCoV2).

MD simulations also revealed that the binding mode of the inhibitor GRL-0617 is completely preserved between SCoV-PLpro and SCoV2-PLpro. As inhibiting PLpro is a promising double-hit therapeutic strategy against COVID-19, GRL-0617 can be a potential lead compound for further studies. Until now, several drugs against COVID-19 are approved in the European Union, but there was no small molecule drug targeting SCoV2-PLpro among them. Ongoing research on further development of GRL-0617 derivatives and additional potential PLpro drugs is encouraging [Osipiuk et al., 2021, Shen et al., 2021, Welker et al., 2021]. Further, research teams from New York and the University of Texas discovered that the oral agents simeprevir, paritaprevir and vaniprevir, which are approved for hepatitis C therapy, also block PLpro and suppress viral replication in cell culture when administered in synergy with remdesivir, increasing remdesivir's antiviral activity 10-fold [Bafna et al., 2021].

Difference in phosphorylation kinetics and conformational changes of casein kinase 1

8.1 Introduction

Casein kinases 1 (CK1) belong to the serine/threonine-selective enzymes and play essential roles in various cellular functions. They are involved in the regulation of DNA repair, cytokinesis, signal transduction pathways and the circadian rhythm [Knippschild et al., 2005]. Seven distinct genes encoding the mammalian CK1 isoforms α, β, γ1-3, δ and ε have been characterized. To date, α, δ and ε are known to be important regulators, but the role of the γ isoforms is not very well understood [Fulcher and Sapkota, 2020].

One role of CK1δ is to control the oocyte quality by phosphorylating the protein TAp63α. TAp63α is present in high concentrations in the oocytes, acting as a quality control factor, and can cause infertility in women after chemotherapy. When DNA is damaged, oocytes initiate programmed cell death in order to minimize genetic defects in the offspring. This process, also known as apoptosis, is regulated in oocytes by TAp63α [Tuppi et al., 2018]. TAp63α is present in an inactive dimeric conformation in non-damaged oocytes. DNA damages from radio- or chemotherapy cause TAp63α to be phosphorylated, which converts TAp63α to the active tetrameric conformation, resulting in apoptosis (Fig. 8.1). As a result, female patients receiving chemotherapy potentially enter menopause much earlier. Premature death of oocytes is not only associated with infertility, but also with other hormone-dependent problems such as osteoporosis. A potential therapeutic strategy to prevent TAp63α phosphorylation leading to apoptosis is to inhibit CK1δ. Indeed, when CK1δ in mice was blocked by inhibitors, their oocytes remained intact when exposed to chemotherapeutic agents [Gebel et al., 2020].

To initiate oocyte apoptosis, TAp63α needs to be sequentially phosphorylated by two kinases, CHK2 and CK1δ. A single phosphorylation by CHK2 is not sufficient to trigger tetramerization, but CK1δ is required to break the auto-inhibitory complex by attaching another four phosphates in a +3 pattern to the primed TAp63α. Upon phosphorylation by CK1δ, the dimeric state of TAp63α is converted into the active tetrameric state due to charge repulsion [Coutandin et al., 2016]. While the first two CK1 phosphorylation events are fast, the third one is slow, which is the decisive step to form the active conformation. The tetramerization is unidirectional, and the third phosphorylation by CK1δ is the point of no return [Gebel et al., 2020]. Hence, the third phosphorylation event determines the threshold of DNA damage necessary for induction of apoptosis. Here, I studied the underlying structural mechanism for the difference in the kinetic behavior of the phosphorylation
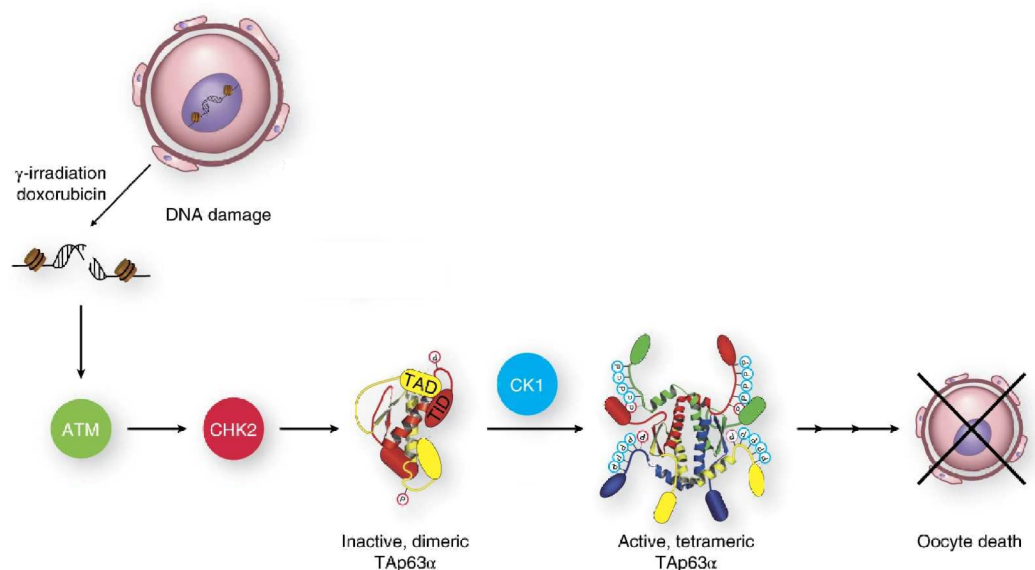
Figure 8.1: **Schematic representation of the TAp63α dimer-to-tetramer activation cascade in oocytes.** Adapted from Tuppi et al. [2018], with kind permission by Springer Nature.

events with MD simulations and found evidence for on an unusual CK1δ/TAp63α substrate interaction.

CK1 kinases can be regulated by autophosphorylation, which leads to inhibition of kinase activity. The C-terminal tails of CK1 family, varying in sequence and length, can serve as substrates of autophosphorylation [Cegielska et al., 1998, Gietzen and Virshup, 1999, Graves and Roach, 1995, Hoekstra et al., 1994]. Autophosphorylation within the kinase domain of CK1ε has also been observed [Cegielska et al., 1998], but a regulatory effect on activity or cellular function has not been found. More recently, our collaborators determined a conserved mechanism of CK1 regulation involving autophosphorylation of the kinase domain [Cullati et al., 2022]. Autophosphorylation of a threonine (Thr220 in human CK1δ), located at the N-terminus of the αG helix in proximity to the substrate binding site, reduces kinase activity. Here, I investigated the effect from phosphorylation with MD simulations and discovered that plasticity of αG is increased upon autophosphorylation of Thr220.

In the following, the three-letter code is used for the kinase and the single amino acid letter code for the TAp63α peptide.

8.2 Methods

8.2.1 MD simulations

In total, seven different MD simulation set-ups were performed. For three simulation set-ups, the X-ray crystal structure from the PDB entry 6ru8 [Gebel et al., 2020] served as starting point. The triple phosphorylated peptide was once elongated to TPpSSAp-STVpSVGSSETRGER with charged termini (S592 and E593 coordinates from PDB entry 6ru7), and once elongated to the shorter peptide ACE-TPpSSApSTVpSVGSSETRG-NME

capped with N-terminal acetyl and C-terminal methylamino capping groups using VMD Molefracture [Humphrey et al., 1996] and Modeller [Šali and Blundell, 1993]. In the third set-up, the point mutation V589A was introduced into the longer peptide using Modeller.

For CK1γ3, two simulations were performed based on X-ray crystal structures of the WT and the Leu257Lys form. Both the X-ray crystal structures of the WT and the mutant are unpublished and were solved by Apirat Chaikuad, a Postdoctoral Scientist in the group of Stefan Knapp, and Jakob Gebel, a Postdoctoral Scientist in the group of Volker Dötsch from the Goethe-University, respectively. In both set-ups, the bound peptides were elongated to ACE-TPpSSApSTVpSVGSSETRG-NME using Modeller.

For CK1δ, three independent MD runs of two additional set-ups each without bound peptide were performed, once in the WT form and once with phosphorylated Thr220 (pThr220) using chain A of PDB entry 4tn6 [Wager et al., 2014] and PDB entry 7p7f [Cullati et al., 2022] as initial structures, respectively.

In all set-ups, all crystallographic water molecules and ions within 10 Å of the protein were retained. Bound ligands were removed, ADP was replaced with ATP and a complexed Mg^{2+} ion was added using PDB entry 1csn [Xu et al., 1995] as a template, by superimposing the protein backbones and aligning the nitrogen atoms of ATP with the crystallographic ADP. Missing side chains were added using the software Modeller. I protonated Asp128 and set up all other ionizable residues in their physiological protonation state. All MD simulations were carried out with Gromacs 2018 [Abraham et al., 2015] using the AMBER99SB*-ILDN-q force field [Hornak et al., 2006a, Best and Hummer, 2009, Best et al., 2012, Lindorff-Larsen et al., 2010], combined with TIP3P water [Jorgensen et al., 1983], 150 mM NaCl [Mamatkulov and Schwierz, 2018], ATP [Meagher et al., 2003], and phosphothreonine or phosphoserine [Homeyer et al., 2006]. After energy minimization, I performed five rounds of equilibration, in which I successively decreased the position restraints on protein heavy atoms, first in an NVT ensemble (0.25 ns) and then in an NPT ensemble (4× 0.5 ns) using a Berendsen thermostat and barostat [Berendsen et al., 1984]. The production runs were run at 310 K in an NPT ensemble using a Nosé–Hoover thermostat [Evans and Holian, 1985, Nosé, 1984]. The pressure was maintained at 1 bar with a Parrinello–Rahman barostat [Parrinello and Rahman, 1981].

8.2.2 Analysis

For the peptide-bound simulations of CK1δ, the minimum heavy-atom distances between E593 and Arg127 and between E593 and Lys154, as well as the minimum heavy-atom distance between V589 or A589 (side chain) and the protein CK1δ were monitored. For the CK1γ3 simulations, the minimum heavy-atom distances between V587 and Leu257, and between V587 and Lys257 were measured. The distances were obtained at 1 ns intervals using the *gmx mindist* tool. The raw distance data were processed using moving average smoothing with a window size of 5 ns. For the replicate simulations of CK1δ in the WT and phosphorylated form, the secondary structure of residues 220-235 was monitored at 10 ns intervals using the *gmx do_ dssp* tool [Kabsch and Sander, 1983].

8.3 Results and discussion

8.3.1 Phosphorylation of the third CK1δ site is the slowest, arising from unusual enzyme-substrate interactions.

Our collaborators could demonstrate that the kinetic behavior of CK1δ phosphorylation sites differ using nuclear magnetic resonance (NMR) based phosphorylation experiments [Gebel et al., 2020]. CK1δ phosphorylates pre-phosphorylated TAp63α four times at S585, S588, S591, and T594. The first two phosphorylation events, pS585 and pS588, occur fast, whereas the phosphorylation for the third phosphorylation site, pS591, is 40 times slower. Hence, the third phosphorylation step is the rate-limiting step and in addition the decisive step for activation. The phosphorylation of all sites takes pace via a distributive mode, because the concentration of the double and triple phosphorylated peptide increased beyond the concentration of the kinase. The first two CK1 phosphorylation events have a distributive mode with a very fast product releasing and substrate re-binding kinetics. In contrast, the third CK1δ phosphorylation takes place slowly, implying that pS585 and pS588 delay phosphorylation of the downstream sites S591 and T594. Mutational studies showed that the point mutations S592V and E593G in TAp63α resulted in stark reduction in the difference between the fast and slow phosphorylation kinetics. To understand this observation and the kinetic difference in WT, I performed MD simulations of CK1δ with the triple phosphorylated TAp63α peptide, in WT and with *in silico* mutated V589A. Interestingly, the X-ray crystal structure is in the product-bound state of the peptide (PDB entry: 6ru8). The 3rd phosphorylated residue is located in the active site, and the phosphate group of pS588 is orientated towards the ATP binding site (Fig 1.5a). Crystallization of the product-bound state indicates that the interaction between enzyme and triple phosphorylated peptide is strong, decelerating the subsequent phosphorylation step.

MD simulations revealed the stabilizing interactions of CK1δ and the triple-phosphorylated TAp63α. Both WT simulations differing in the length of the peptide demonstrated strong electrostatic interactions between the positively charged side chains Arg127, Lys154, and Lys171 in CK1δ with E593 of TAp63α. The set-ups were modelled with either a long peptide construct with charged termini (TPpSSApSTVpSVGSSETRGER) or a shorter peptide (ACE-TPpSSApSTVpSVGSSETRG-NME) capped with N-terminal acetyl and C-terminal methylamino capping groups. The long peptide sequence is the same as used in the experimental kinetic measurements. The short peptide mimics the full-length protein, because E597 and R598 may not be accessible to CK1. Both simulations consistently displayed strong interactions of E593 with a basic cluster in CK1δ consisting of Arg127, Lys154, and Lys171. For the longer construct, E593 forms a persistent salt bridge with Arg127 (distance < 3 Å), and binds transiently to Lys154 (Fig. 8.2a,b). Similar results are observed for the shorter TAp63α construct (Fig. 8.3). In addition, MD simulations highlighted the importance of the van-der-Waals interactions between V589 and CK1δ. Comparison of the two 1 μs long simulations of the WT and the V589A peptide showed that the V589A mutation weakened the hydrophobic interactions with CK1δ and increased the flexibility of the peptide in this region (Fig. 8.2c,d). In the WT set-up, V589 remained
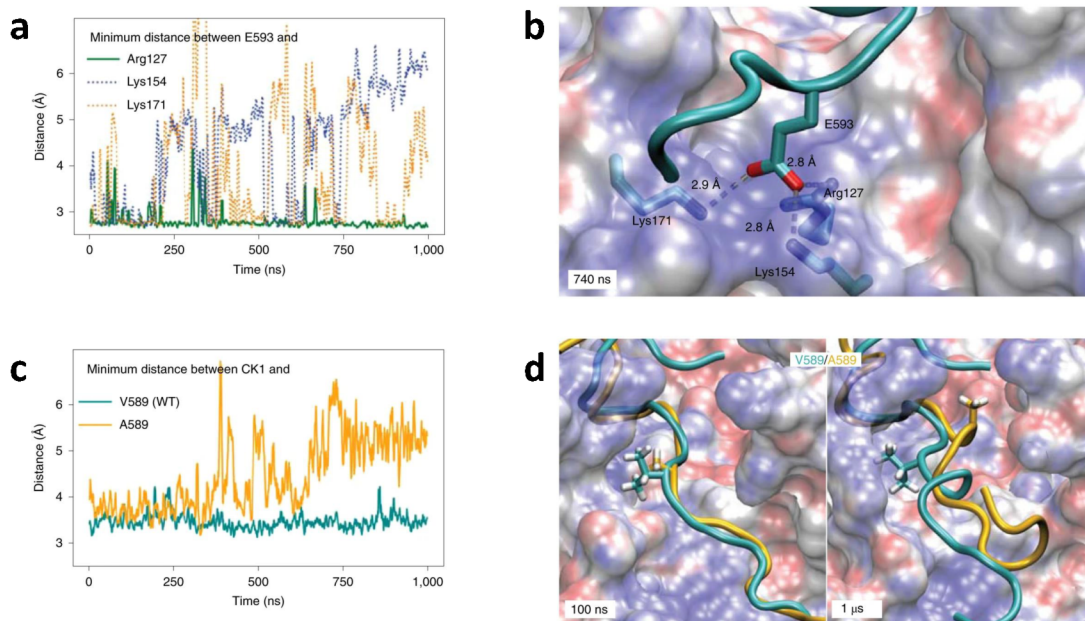
Figure 8.2: **MD simulations of CK1δ in complex with TAp63α revealed persistent electrostatic and hydrophobic interactions.** Reprinted from Gebel et al. [2020], with kind permission by Springer Nature. **a,** Minimum heavy-atom distances of E593 to Arg127, Lys154, and Lys171. **b,** Representative snapshot at 740 ns, zooming in on the C-terminal region of the p63 peptide. CK1 is shown as a transparent electrostatic surface (blue/red for positive/negative charge), and the p63 peptide is represented as a cyan cartoon. Residues E593, Arg127 and Lys154 are highlighted. The minimum distances between E593 and the basic residues are indicated. **c,** Minimum heavy-atom distances of CK1 to V589 and A589 in wild type and mutant p63 peptides are shown as a function of time. **d,** Snapshots at 100 ns and 1 µs, zooming in on the C-terminal region of the p63 peptide. CK1 structures are superimposed. CK1 from the WT simulation is represented as a transparent electrostatic surface (blue/red for positive/negative charge), and the p63 peptide is shown in cartoon representation (cyan, wild type; yellow, V589A) with V589 and A589 highlighted. Results are shown for the longer TAp63α peptide.

adhered to the CK1 protein surface, whereas A589 became solvated and moved away from the CK1 protein surface. This result agrees well with the experimental mutational analysis, which demonstrated that a large hydrophobic residue in the i+1-position is required to slow down phosphorylation at the third CK1 site.

With MD simulations, I could trace the slow kinetics of the third phosphorylation step, which is the point of no return, to on an unusual enzyme-substrate interaction. The simulations identified the stabilizing interactions between CK1δ and TAp63α as persistent salt-bridges and tight hydrophobic contacts, in a form unfavorable for phospho-transfer. Consistently, experimental mutational and MD studies demonstrated that E593 and V589 of the TAp63α peptide are key residues for interaction with CK1δ.

8.3.2 Phosphorylation of the third CK1γ site is non-existent, potentially resulting from point mutation.

For the γ3 isoform, the phosphorylation kinetics of TAp63α are different. Here, the phosphorylation stalls after the second phosphorylation site and the third phosphorylation does not take place [Gebel, 2022]. Structural comparison between the γ3 and δ isoform shows
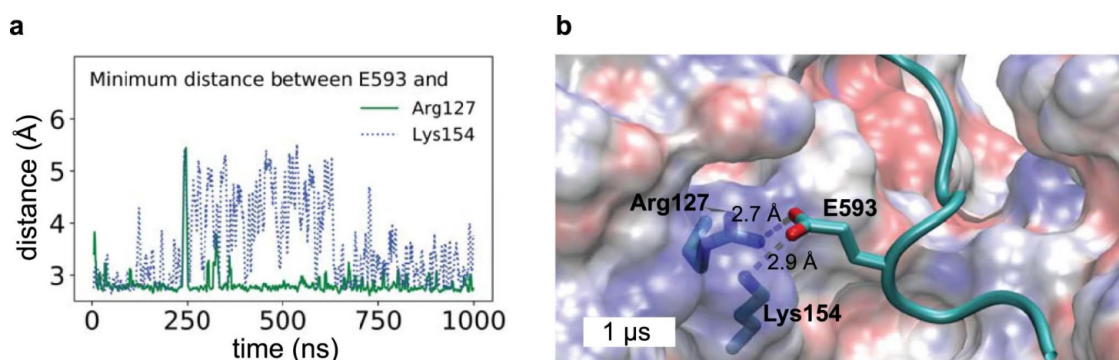
Figure 8.3: **MD simulation of CK1δ in complex with a shorter TAp63α peptide (ACE-TPpSSApSTVpSVGSSETRG-NME) with N-terminal acetyl and C-terminal methylamino capping groups revealed similar results as the longer peptide.** Reprinted from Gebel et al. [2020], with kind permission by Springer Nature. **a,** Minimum heavy-atom distances of E593 to Arg127 and Lys154 are shown as a function of time. **b,** Snapshot at 1 μs, zooming in on the C-terminal region of the shorter p63 peptide. CK1 is shown as a transparent electrostatic surface (blue/red for positive/negative charge) and the p63 peptide is represented as a cyan cartoon. The residues E593, Arg127 and Lys154 are highlighted. The minimum distances between E593 and the basic residues are indicated.

no large differences (backbone RMSD of kinases: 1.8 Å). However, two side chains next to the binding cleft are different: Lys75 and Lys221 in the δ isoform correspond to Arg79 and Leu257 in the γ3 isoform. Indeed, NMR based phosphorylation experiments showed that the single mutation Lys221Leu in the CK1δ form could recreate the γ3 behavior displaying only two phosphorylation events. The reversed mutation in the γ3 isoform, Leu257Lys, results in an effectively non-functional kinase, which, however, has a similar behavior to the δ isoform at a very low level in terms of kinetics. The X-ray crystal structure of the CK1γ3 mutant Leu257Lys did not reveal any structural changes compared with the WT form. Comparison of the WT and mutant form with MD simulations revealed that the interactions between CK1 and TAp63α differ (Fig. 8.4). In the WT form, Leu257 of CK1 strongly interacts with V587 of the TAp63α peptide. This hydrophobic interaction is lost in the mutated form. Within 10 μs Leu257 in the WT form remained attached to V587, whereas Lys257 and V587 did not interact and were separated by water molecules (Fig. 8.4b,c). The strong interaction of Leu257 and V587 is potentially one key factor for eliminating the third phosphorylation step in the γ3 isoform.

8.3.3 pThr220 destabilizes N-terminal part of αG segment in apo CK1δ

Cullati et al. [2022] demonstrated that autophosphorylation of CK1 at the Thr220 position results in enzyme inhibition. The first solved X-ray crystal structure (PDB entry: 7p7f) with the phosphorylated Thr220 did not exhibit any structural changes compared to the WT form (backbone RMSD of αG helix: 1.4 Å after backbone protein alignment). The phosphorylatable threonine is located at the end of loop FG preceding the αG helix. To understand why the phosphorylated form inhibits activity, I performed MD simulations of both the WT and the phosphorylated form. I found that the WT αG segment in apo CK1δ can adopt helix and loop conformation (Fig. 8.5a (left image),b). This finding is consistent
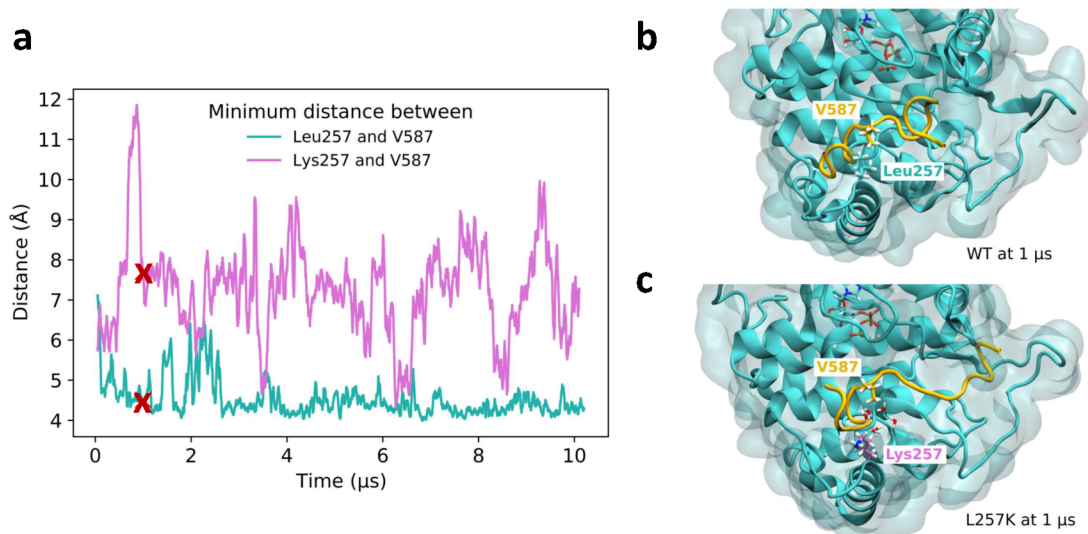
Figure 8.4: **MD simulations of CK1γ3 in complex with TAp63α indicated the impact of Leu257 on the phosphorylation kinetics. a,** Minimum heavy-atom distances of V587 to Leu257 and Lys257 are shown as a function of time. Crosses denote time frames of snaphots at 1 μs. **b,** Snapshot of WT form at 1 μs with highlighted Leu221 and V587 residues. **c,** Snapshot of L257K form at 1 μs with highlighted Lys221 and V587 residues. Water molecules located between Lys221 and V587 are shown with atomic detail.

with the high B-factors of this segment, which indicate high flexibility. The plastic nature of this segment could affect kinase-substrate interactions [Philpott et al., 2020, Shinohara et al., 2017]

However, the rate of transitioning between different states of the αG segment differed significantly between the WT and phosphorylated form (Fig. 8.5a). For the native form, I observed two different scenarios: (1) a stable helical structure and (2) a complete distortion of the helix throughout the 2 μs time frame. Contrary to the native form, the simulations of pThr220 revealed a greater degree of flexibility in the N-terminal part of the αG helix where pThr220 is situated, whereas the helical structure at the C-terminal end remained more stable. A second X-ray crystal structure of the phosphorylated form, which was solved in the anion-free condition, showed αG in a partially unwound conformation (PDB entry: 7p7g), supporting our findings. The destabilized segment observed in the X-ray crystal structure nicely overlays with the predicted region from MD simulations (Fig. 8.5c). The structural changes can alter the conformation of the substrate binding cleft, resulting in reduced activity. Hence, autophosphorylation can regulate the CK1 enzyme activity.

A Bayesian model comparison was performed as follows[1] to assess the statistical significance of the finding that phosphorylation destabilizes the αG helix: The unfolding of the N-terminal helix was observed in several simulations (Table 8.1). Because refolding events were not observed, we focus on the rates of helix unfolding of WT and pThr220. In the following, we assume that the unfolding process follows a first-order kinetic model, helix to coil ($H \rightarrow C$), with rates k and k' for the WT and pThr220 forms, respectively.

---

[1] Analysis by G. Hummer.
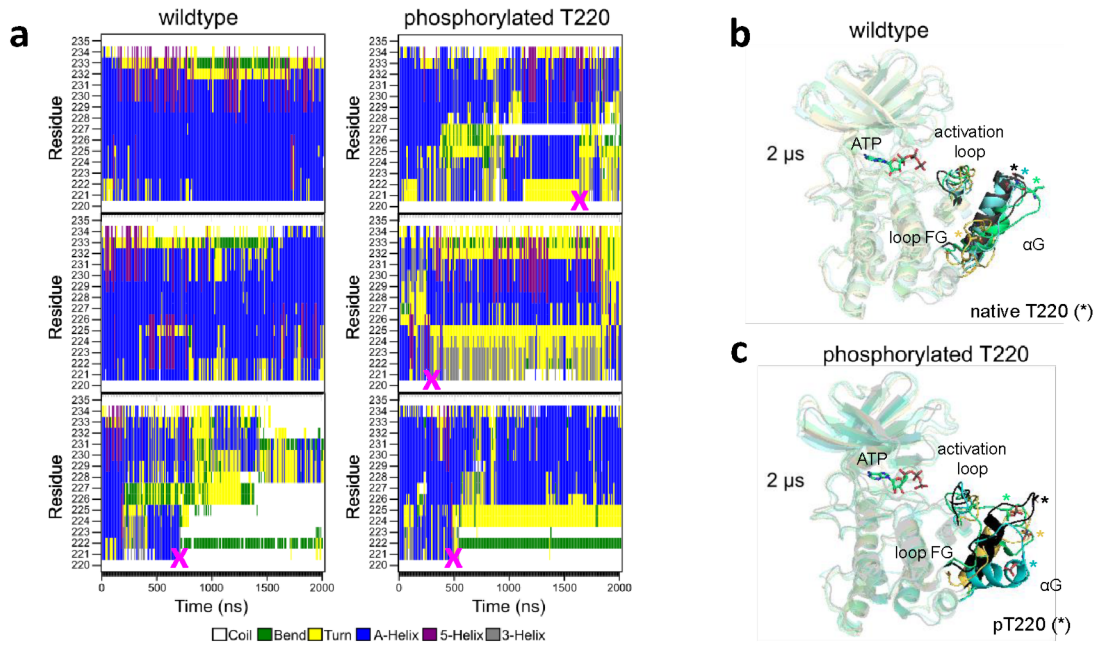
Figure 8.5: **pThr220 destabilizes N-terminal part of αG helix.** Reprinted from Cullati et al. [2022], with kind permission by Elsevier. **a,** The secondary structure of αG residues for the unphosphorylated (left) and pThr220 (right) kinase obtained from 3× 2 µs MD simulations. Time points when the N-terminal part of the helix unfolded are marked by magenta crosses. **b,** Conformations at 2 µs from three replicates for unphosphorylated form (green, cyan, and yellow) superpositioned with the WT X-ray crystal structure (PDB entry: 4tn6) shown in black. Asterisk shows the positions of Thr220. **c,** Conformations at 2 µs from three replicates for phosphorylated form (green, cyan, and yellow) superpositioned with the pThr220 X-ray crystal structure solved in the anion-free condition (PDB entry: 7p7g) shown in black. Asterisk shows the positions of pThr220.
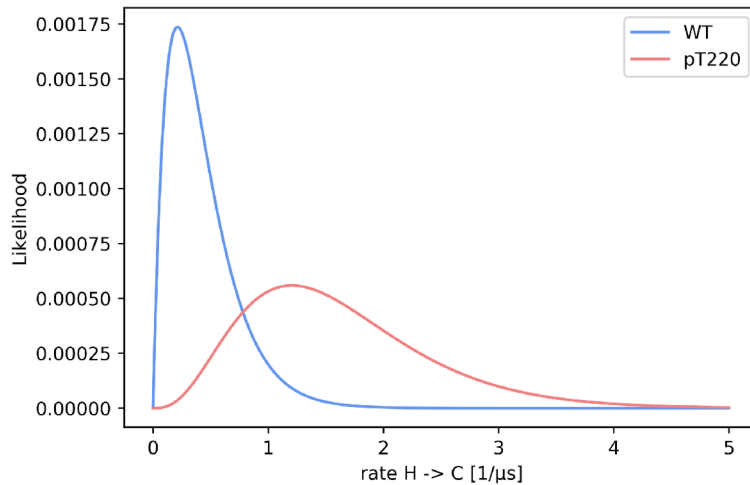
Figure 8.6: **Likelihoods of unfolding rates for WT and phosphorylated form.**

Table 8.1: Statistics of unfolding events in WT and pThr220 MD simulations. Listed are time points of unfolding events (event = 1) or, if no unfolding occurred (event = 0), of the run duration.

run	t_i (WT) [µs]	event	τ_i (pThr220) [µs]	event
1	2.00	0	1.64	1
2	0.72	1	0.50	1
3	2.00	0	0.35	1

Given the event statistics in Table 8.1, the likelihoods then are

$$L_{WT} = ke^{-k(t_1+t_2+t_3)} = ke^{-kt} \tag{8.1}$$

with $t = t_1 + t_2 + t_3$ the cumulative time in the helical state for WT, and

$$L_{pThr220} = (k')^3 e^{-k'(\tau_1+\tau_2+\tau_3)} = (k')^3 e^{-k'\tau} \tag{8.2}$$

with $\tau = \tau_1 + \tau_2 + \tau_3$ the cumulative time in the helical state for pThr220 (Fig. 8.6). For a Jeffreys prior, prior$(k) \propto 1/k$ for $k > 0$, and analogously for k', the normalized posteriors become

$$p_{WT}(k|\text{data}_{WT}) = te^{-kt} \tag{8.3}$$

and

$$p_{pThr220}(k'|\text{data}_{pThr220}) = \frac{1}{2}\tau^3(k')^2 e^{-k'\tau} \tag{8.4}$$

We now consider two mutually exclusive models, $k \geq k'$ and $k < k'$. A priori, we consider both models to be equally likely. Then, given the data in Table 8.1, the evidence for model 1 with $k \geq k'$ is

$$
\begin{aligned}
&p(k \geq k'|\text{data}_{WT}, \text{data}_{pThr220}) \\
&= \int_0^\infty dk \int_0^k dk' p_{WT}(k|\text{data}_{WT})\, p_{pThr220}(k'|\text{data}_{pThr220}) \\
&= \frac{1}{(1+t/\tau)^3} \approx 0.041
\end{aligned}
\tag{8.5}
$$

Conversely, the evidence for model 2 with $k < k'$ is

$$
\begin{aligned}
&p(k < k'|\text{data}_{WT}, \text{data}_{pThr220}) \\
&= 1 - p(k \geq k'|\text{data}_{WT}, \text{data}_{pThr220}) \approx 0.959
\end{aligned}
\tag{8.6}
$$

Given equal a priori probabilities of the two models, the Bayes factor in favor of model 2 over model 1 is

$$K = \frac{p(k < k'|\text{data}_{WT}, \text{data}_{pThr220})}{p(k \geq k'|\text{data}_{WT}, \text{data}_{pThr220})} \approx 23.3 \tag{8.7}$$

The cumulative evidence from 2×3 MD simulations therefore strongly favors a model in which phosphorylation of Thr220 accelerates the rate of helix unfolding.

Hence, we could show that the phosphorylated form tends to destabilize the N-terminal part of the αG helix. Visually inspecting all three trajectories of the phosphorylated form revealed that the negatively charged phosphate group interacts with one or two arginines, in particular Arg222 and Arg227, which leads to the deformation of the helix structure (Fig. 8.7). In addition to the strong electrostatic interactions, the change in the helix capping and dipolar moment of the added phosphate group potentially weakens the helical structure.
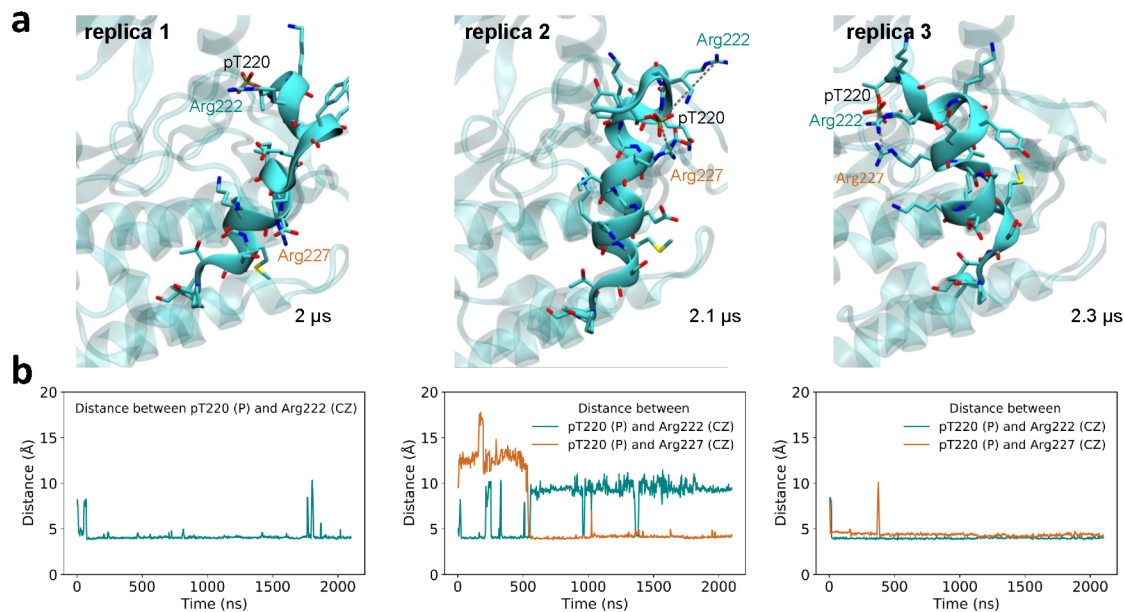
Figure 8.7: **N-terminal αG helix unfolding due to pThr220-Arg interactions. a,** Snapshots for all three phosphorylated simulation replicates with αG helix residues highlighted with atomic detail. **b,** Distances between the phosphate of pThr220 and the terminal carbon atom of the Arg side chains of both Arg222 and Arg227 are shown as a function of time.

8.4 Summary and conclusion

CK1 kinases regulate a variety of important cellular pathways, including DNA repair. In oocytes, CK1 is associated with activating a cell-death program that leads to infertility in women after chemotherapy-induced DNA damage. A key step in this process is the third phosphorylation of TAp63α, which converts TAp63α into an active conformation. With MD simulations, I could trace the slow kinetics of this decisive step to on an unusual enzyme-substrate interaction, compared to the other three phosphorylation events. Using MD simulations, I identified the stabilizing interactions between CK1δ and TAp63α as persistent salt bridges and tight hydrophobic contacts in a form unfavorable for phospho-transfer. In contrast, the phosphorylation by the CK1γ3 isoform is stalled after the second phosphorylation and the third phosphorylation step is eliminated. In our simulations, I could identify a potential key factor for this behavior: V587 of the TAp63α peptide is attached via hydrophobic interactions to Leu257 in the γ3 isoform, which corresponds to Lys257 in the δ isoform. In addition, our experimental collaborators found that CK1 can be autophosphorylated, resulting in reduced enzyme activity. MD simulations showed that the phosphorylated form exhibited greater plasticity than the native form. In particular, the integrity of the substrate binding site in the phosphorylated form was disrupted, which explains the reduced activity. Thus, autophosphorylation at Thr220 potentially controls the conformation and integrity of the substrate binding pocket.

Chapter 9

Conclusions

I used physics-based computational methods to study the flexibility of binding sites and enzyme-substrate interactions. Investigating ligand-induced conformational dynamics and flexibility of a protein are of high interest to advance the computational contribution to drug discovery studies [Adelusi et al., 2022]. This can enable and accelerate the drug design process. Proteins are dynamic biomacromolecules that can have diverse and nearly isoenergetic conformational states. Ligand binding can shift the equilibrium of this conformational ensemble. Cryptic pockets are binding sites in proteins that become apparent only in ligand-bound structures, and are not visible in unbound structures. These sites provide an opportunity to target proteins that were previously considered undruggable. However, their hidden nature makes it difficult to identify and exploit them. We investigated protein conformational changes upon ligand binding in six different systems. In all systems, it is known from experiment that a specific group of ligands targeting the binding site can induce conformational changes. For FVIIa, the observed conformational change is the distortion of a β-sheet structure in the S1 pocket, while for renin a non-functional flap opens up combined with Trp39 side chain rotation. Four additional drug targets from the dataset CryptoSite show displaceable shallowly buried aromatic residues: NPC2, p38α, ricin, and Eg5.

For all systems, except for renin, computational methods could have predicted these conformational changes. For FVIIa, the conformational change can be captured with unbiased MD simulations. I found that the S1 pocket flexibility is not only present in FVIIa, but that it is transferable to other serine peptidases including thrombin, matriptase, and tryptase. For renin, the full pocket opening could not be sampled, even with established biasing protocols such as adding organic fragments to the simulation box and increasing the water-protein interaction using the SWISH technique. The biasing protocols enhanced the protein flexibility, but did not disrupt the hydrophobic interactions between Trp39 and Phe112. Well-tempered metadynamics simulation indicated a high free energy barrier for the Trp39 side chain rotation ($\approx$35 kJ/mol), which explains the sampling difficulties. The simulations of renin suggest that small molecules can induce major conformational changes yet achieve high-affinity binding. For all four proteins from the CrypstoSite data set, I could resolve the opening of the cryptic pockets. By using different simulations methods, we gained insights into the respective binding mechanisms and place them in the spectrum of induced fit and conformational selection. I used three different methods to probe and induce the opening of cryptic pockets starting from the unbound structures: unbiased MD simulations, mixed-solvent MD simulations with 1 M benzene molecules, and

the enhanced sampling method SWISH. In mixed-solvent MD simulations of NPC2, two benzene molecules sequentially entered into the cryptic pocket. Opening of the deep pocket was associated with ligand binding, which advocates for a major role of induced fit in the ligand binding mechanism. By contrast, the cryptic pocket of p38α was flexible already in the unbiased MD simulation. In mixed-solvent MD simulations of p38α, a benzene molecule inserted into the cryptic pocket, indicating a balance of induced fit and conformational selection. In contrast to NPC2 and p38α, the conformational changes of the cryptic pockets of ricin and Eg5 were largely independent of the presence of probe molecules, suggesting that conformational selection dominates the ligand binding mechanism. Interestingly, benzene binding in ricin and Eg5 revealed several additional binding pockets besides the known cryptic sites. In conclusion, mixed-solvent simulations with 1 M benzene molecules were able to expose cryptic sites in all systems, except for renin. For renin, only the flap movement was captured, but not the full pocket opening. If the respective energy barrier is too high, i.e., above 35 kJ/mol, the existing methods fail. In this case, protein surfaces can potentially be scanned for shallowly buried aromatic residues, which can be biased specifically to assess their flexibility. In general, our simulations show that the opening mechanisms of cryptic pockets are specific to the system under consideration. Hence, establishing a robust method that works on all systems remains to be considerably challenging. However, I also found that mixed-solvent simulation perform well on four out of five studied dissimilar cryptic sites, ranging from conformational selection to induced-fit. This makes mixed-solvent simulation an efficient method for assessing druggability.

Furthermore, I used MD simulations to study enzyme-substrate interactions in two drug targets: the PLpro of SARS-CoV-2 and the human CK1. PLpro plays a critical role in coronavirus replication. In addition, PLpro can suppress the innate immune response by preferentially cleaving ISG15 compared with ubiquitin. With MD simulations, I could confirm that PLpro from SARS-CoV-2 interacts more tightly with ISG15. While ISG15 remained bound in three independent MD runs of 3.2 μs, the distal ubiquitin of di-ubiquitin separated from PLpro in four out of six runs on a microsecond timescale. I observed a water-mediated dissociation mechanism for ubiquitin and identified L75T of PLpro as a key mutation distinguishing the earlier SARS-CoV and new SARS-CoV-2 coronavirus, as it weakens the hydrophobic cluster within the binding interface. Hence, pharmacological inhibition of PLpro in SARS-CoV-2 not only blocks viral replication, but also simultaneously boosts the antiviral immune response.

CK1 regulates a variety of important cellular pathways, including DNA repair. After chemotherapy-induced DNA damage, CK1 is associated with activating a cell-death program in oocytes that leads to infertility in women. A key step in this process is the third phosphorylation of TAp63α, which converts TAp63α into an active conformation. With MD simulations, I could trace the slow kinetics of this decisive step—compared to the other three phosphorylation events—to on an unusual enzyme-substrate interaction. The simulations identified the stabilizing interactions between CK1 and TAp63α as persistent salt bridges and tight hydrophobic contacts in a form unfavorable for phospho-transfer. When inhibiting CK1 in mice, the oocytes remained intact, even under the influence of

chemotherapeutic agents. Our experimental collaborators found that CK1 can be autophosphorylated, resulting in reduced enzyme activity. Using MD simulations, I observed that the phosphorylated form exhibited greater plasticity than the non-phosphorylated form. In particular, the integrity of the substrate binding site in the phosphorylated form was altered, which explains the reduced activity.

In summary, MD simulations allowed us to investigate binding site flexibility and enzyme-substrate interaction in atomic detail. We could contribute to the assessment of cryptic pockets in specific drug targets and to the understanding of disease mechanisms, in particular COVID-19 and infertility in women after chemotherapy.

Advances in computational resources will push the time boundaries of simulations further and will allow for more sampling time both for investigating cryptic sites and enzyme-substrate interactions. Shan et al. [2022] recently used extensive computational resources to decipher ligand binding to a cryptic site. Current approaches will be developed further, for example a different procedure of a mixed-solvent MD simulation scheme was delevoped by Ung et al. [2016], Lal Gupta and Carlson [2022], which uses only miscible, organic solvent, such as pyrimidine. By using solely miscible fragments, no inter-ligand repulsion terms are necessary to prevent their aggregation. However, in this case, the choice of fragments is limited. At present, mixed-solvent simulations are already routinely used for target assessments in the pharmaceutical industry. With rising computational power, mixed-solvent MD simulations will not only support target assessment, but also help in lead generation by improving activity and reducing the toxicity of compounds in early drug discovery research. One simulation of a desirable target protein in a mixed-solvent with several diverse fragments can in principle construct the potential interaction pattern in the binding site. This can inform drug designers which binding site residues can be targeted to optimize compounds further. Additionally, current methodologies for sampling of cryptic sites can be improved by combining MD with fragment docking and machine learning approaches [Vajda et al., 2018]. In general, the combination of artificial intelligence and physics-based methods will shape the future of computational drug design.

Bibliography

Annex: War weapons list. `https://germanlawarchive.iuscomp.org/?p=743`. Accessed: 2022-05-25.

Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, 2015. ISSN 23527110. doi: 10.1016/j.softx.2015.06.001.

Temitope Isaac Adelusi, Abdul-Quddus Kehinde Oyedele, Ibrahim Damilare Boyenle, Abdeen Tunde Ogunlana, Rofiat Oluwabusola Adeyemi, Chiamaka Divine Ukachi, Mukhtar Oluwaseun Idris, Olamide Tosin Olaoba, Ibrahim Olaide Adedotun, Oladipo Elijah Kolawole, et al. Molecular modeling in drug discovery. *Informatics in Medicine Unlocked*, page 100880, 2022.

Hans C Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384–2393, 1980.

Natalia S Andreeva and Lev D Rumsh. Analysis of crystal structures of aspartic proteinases: On the role of amino acid residues adjacent to the catalytic site of pepsin-like enzymes. *Protein Science*, 10(12):2439–2450, 2001.

Khushboo Bafna, Kris White, Balasubramanian Harish, Romel Rosales, Theresa A Ramelot, Thomas B Acton, Elena Moreno, Thomas Kehrer, Lisa Miorin, Catherine A Royer, et al. Hepatitis C virus drugs that inhibit SARS-CoV-2 papain-like protease synergize with remdesivir to suppress viral replication in cell culture. *Cell Reports*, 35(7):109133, 2021.

Alaji Bah, Laura C Garvey, Jingping Ge, and Enrico Di Cera. Rapid kinetics of Na^+ binding to thrombin. *Journal of Biological Chemistry*, 281(52):40049–40056, 2006. ISSN 0021-9258. doi: 10.1074/JBC.M608600200. URL `https://www.sciencedirect.com/science/article/pii/S0021925820768407`.

S. Paul Bajaj, Amy E. Schmidt, Sayeh Agah, Madhu S. Bajaj, and Kaillathe Padmanabhan. High resolution structures of p-aminobenzamidine- and benzamidine-VIIa/ soluble tissue factor: Unpredicted conformation of the 192-193 peptide bond and mapping of Ca^{2+}, Mg^{2+}, Na^+, and Zn^{2+} sites in factor VIIa. *Journal of Biological Chemistry*, 281(34): 24873–24888, 2006. ISSN 00219258. doi: 10.1074/jbc.M509971200. URL `https://pubmed.ncbi.nlm.nih.gov/16757484/`.

David W. Banner, Allan D'Arcy, Christiana Chène, Fritz K. Winkler, Arabinda Guha, William H. Konigsberg, Yale Nemerson, and Daniel Kirchhofer. The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature*, 380

(6569):41–46, 1996. ISSN 0028-0836. doi: 10.1038/380041a0. URL http://www.nature.com/articles/380041a0.

Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Physical Review Letters*, 100(2):020603, 2008.

Kristen M Bartoli, Jelena Jakovljevic, John L Woolford Jr, and William S Saunders. Kinesin molecular motor Eg5 functions during polypeptide synthesis. *Molecular Biology of the Cell*, 22(18):3420–3430, 2011.

Christopher I Bayly, Piotr Cieplak, Wendy Cornell, and Peter A Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry*, 97(40):10269–10280, 1993.

Dmitri Beglov, David R Hall, Amanda E Wakefield, Lingqi Luo, Karen N Allen, Dima Kozakov, Adrian Whitty, and Sandor Vajda. Exploring the structural origins of cryptic sites on proteins. *Proceedings of the National Academy of Sciences*, 115(15):E3416–E3425, 2018.

Miklós Békés, Reggy Ekkebus, Huib Ovaa, Tony T Huang, Christopher D Lima, et al. Recognition of Lys48-linked di-ubiquitin and deubiquitinating activities of the SARS coronavirus papain-like protease. *Molecular Cell*, 62(4):572–585, 2016.

H. J.C. Berendsen, J. P.M. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics*, 81(8):3684–3690, 1984. ISSN 00219606. doi: 10.1063/1.448118. URL http://aip.scitation.org/doi/10.1063/1.448118.

Ondrej Bernatik, Ranjani Sri Ganji, Jacomijn P Dijksterhuis, Peter Konik, Igor Cervenka, Tilman Polonio, Pavel Krejci, Gunnar Schulte, and Vitezslav Bryja. Sequential activation and inactivation of Dishevelled in the Wnt/β-catenin pathway by casein kinases. *Journal of Biological Chemistry*, 286(12):10396–10410, 2011.

Robert B. Best and Gerhard Hummer. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *Journal of Physical Chemistry B*, 113(26): 9004–9015, 2009. ISSN 15206106. doi: 10.1021/jp901540t.

Robert B. Best, David De Sancho, and Jeetain Mittal. Residue-specific α-helix propensities from molecular simulation. *Biophysical Journal*, 102(6):1462–1467, 2012. ISSN 00063495. doi: 10.1016/j.bpj.2012.02.024.

Soumendranath Bhakat and Pär Söderhjelm. Flap dynamics in pepsin-like aspartic proteases: A computational perspective using plasmepsin-II and BACE-1 as model systems. *Journal of Chemical Information and Modeling*, 62(4):914–926, 2022. ISSN 1549960X. doi: 10.1021/acs.jcim.1c00840. URL https://doi.org/10.1021/acs.jcim.1c00840.

Tom Blundell, Bancinyane Lynn Sibanda, and Laurence Pearl. Three-dimensional structure, specificity and catalytic mechanism of renin. *Nature*, 304(5923):273–275, 1983.

Regine S Bohacek, Colin McMartin, and Wayne C Guida. The art and practice of structure-based drug design: a molecular modeling perspective. *Medicinal Research Reviews*, 16(1):3–50, 1996.

Max Born and Robert Oppenheimer. Zur Quantentheorie der Molekeln. *Annalen der Physik*, 389(20):457–484, 1927.

Natércia F Brás, Pedro A Fernandes, and Maria J Ramos. Molecular dynamics studies on both bound and unbound renin protease. *Journal of Biomolecular Structure and Dynamics*, 32(3):351–363, 2014.

Vítezslav Bryja, Gunnar Schulte, Nina Rawal, Alexandra Grahn, and Ernest Arenas. Wnt-5a induces Dishevelled phosphorylation and dopaminergic differentiation via a CK1-dependent mechanism. *Journal of Cell Science*, 120(4):586–595, 2007.

Malcolm S Buchanan, Anthony R Carroll, Deborah Wessling, Michael Jobling, Vicky M Avery, Rohan A Davis, Yunjiang Feng, Yafeng Xue, Linda Oster, Tomas Fex, et al. Clavatadine A, a natural product with selective recognition and irreversible inhibition of factor XIa. *Journal of Medicinal Chemistry*, 51(12):3583–3587, 2008.

Tim DH Bugg. *Introduction to enzyme and coenzyme chemistry*. John Wiley & Sons, 2012.

Giovanni Bussi, Davide Branduardi, et al. Free-energy calculations with metadynamics: Theory and practice. *Reviews in Computational Chemistry*, 28:1–49, 2015.

David Casagolda, Beatriz del Valle-Pérez, Gabriela Valls, Ero Lugilde, Meritxell Vinyoles, Juan Casado-Vela, Guiomar Solanas, Eduard Batlle, Albert B Reynolds, José Ignacio Casal, et al. A p120-catenin–CK1ε complex regulates Wnt signaling. *Journal of Cell Science*, 123(15):2621–2631, 2010.

Andrew Castillo and Monica J Justice. The kinesin related motor protein, Eg5, is essential for maintenance of pre-implantation embryogenesis. *Biochemical and Biophysical Research Communications*, 357(3):694–699, 2007.

Aleksandra Cegielska, Kimberly Fish Gietzen, Ann Rivers, and David M Virshup. Autoinhibition of casein kinase I ε (CKIε) is relieved by protein phosphatases and limited proteolysis. *Journal of Biological Chemistry*, 273(3):1357–1364, 1998.

Jit Kong Cheong and David M Virshup. Casein kinase 1: Complexity in the family. *The International Journal of Biochemistry & Cell Biology*, 43(4):465–469, 2011.

Peter Cimermancic, Patrick Weinkam, T Justin Rettenmaier, Leon Bichmann, Daniel A Keedy, Rahel A Woldeyes, Dina Schneidman-Duhovny, Omar N Demerdash, Julie C Mitchell, James A Wells, et al. CryptoSite: expanding the druggable proteome by

characterization and prediction of cryptic binding sites. *Journal of Molecular Biology*, 428(4):709–719, 2016.

Federico Comitani and Francesco Luigi Gervasio. Exploring cryptic pockets formation in targets of pharmaceutical interest with SWISH. *Journal of Chemical Theory and Computation*, 14(6):3321–3331, 2018.

Wendy D Cornell, Piotr Cieplak, Christopher I Bayly, Ian R Gould, Kenneth M Merz, David M Ferguson, David C Spellmeyer, Thomas Fox, James W Caldwell, and Peter A Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995.

Wendy D Cornell, Piotr Cieplak, Christopher I Bayly, and Peter A Kollman. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *Journal of the American Chemical Society*, 115(21):9620–9631, 2002.

Daniel Coutandin, Christian Osterburg, Ratnesh Kumar Srivastav, Manuela Sumyk, Sebastian Kehrloesser, Jakob Gebel, Marcel Tuppi, Jens Hannewald, Birgit Schäfer, Eidarus Salah, et al. Quality control in oocytes by p63 is based on a spring-loaded activation mechanism on the molecular and cellular level. *Elife*, 5:e13909, 2016.

Peter Csermely, Robin Palotai, and Ruth Nussinov. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Nature Precedings*, pages 1–1, 2010.

Jie Cui, Fang Li, and Zheng-Li Shi. Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, 17(3):181–192, 2019.

Sierra N Cullati, Apirat Chaikuad, Jun-Song Chen, Jakob Gebel, Laura Tesmer, Rezart Zhubi, Jose Navarrete-Perea, Rodrigo X Guillen, Steven P Gygi, Gerhard Hummer, et al. Kinase domain autophosphorylation rewires the activity and substrate specificity of CK1 enzymes. *Molecular Cell*, 2022.

Alan W Sousa Da Silva and Wim F Vranken. ACPYPE-Antechamber python parser interface. *BMC research notes*, 5(1):1–8, 2012.

Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An N log (N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.

Xavier Daura, Karl Gademann, Bernhard Jaun, Dieter Seebach, Wilfred F Van Gunsteren, and Alan E Mark. Peptide folding: when simulation meets experiment. *Angewandte Chemie International Edition*, 38(1-2):236–240, 1999.

Beatriz del Valle-Pérez, Oriol Arqués, Meritxell Vinyoles, Antonio García de Herreros, and Mireia Duñach. Coordinated action of CK1 isoforms in canonical Wnt signaling. *Molecular and Cellular Biology*, 31(14):2877–2888, 2011.

Alexey Dementiev, Abel Silva, Calvin Yee, Zhe Li, Michael T Flavin, Hing Sham, and James R Partridge. Structures of human plasma β–factor XIIa cocrystallized with potent inhibitors. *Blood Advances*, 2(5):549–558, 2018.

V Dhanaraj, CG Dealwis, C Frazao, M Badasso, BL Sibanda, IJ Tickle, JB Cooper, HPC Driessen, M Newman, C Aguilar, et al. X-ray analyses of peptide-inhibitor complexes define the structural basis of specificity for human and mouse renins. *Nature*, 357(6378): 466–472, 1992.

Namrita Dhillon and Merl F Hoekstra. Characterization of two protein kinases from Schizosaccharomyces pombe involved in the regulation of DNA repair. *The EMBO journal*, 13(12):2777–2788, 1994.

Enrico Di Cera. Thrombin. *Molecular Aspects of Medicine*, 29(4):203–254, 2008. ISSN 0098-2997. doi: 10.1016/J.MAM.2008.01.001. URL `https://www.sciencedirect.com/science/article/abs/pii/S0098299708000022`.

C D Dickinson, C R Kelly, and W Ruf. Identification of surface residues mediating tissue factor binding and catalytic function of the serine protease factor VIIa. *Proceedings of the National Academy of Sciences of the United States of America*, 93(25):14379–84, 1996. ISSN 0027-8424. doi: 10.1073/pnas.93.25.14379. URL `http://www.ncbi.nlm.nih.gov/pubmed/8962059`.

Christopher M Dobson et al. Chemical space and biology. *Nature*, 432(7019):824–828, 2004.

Christian Drosten, Stephan Günther, Wolfgang Preiser, Sylvie Van Der Werf, Hans-Reinhard Brodt, Stephan Becker, Holger Rabenau, Marcus Panning, Larissa Kolesnikova, Ron AM Fouchier, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *New England Journal of Medicine*, 348(20):1967–1976, 2003.

Jacob D Durrant and J Andrew McCammon. Molecular dynamics simulations and drug discovery. *BMC biology*, 9(1):1–9, 2011.

Charles Eigenbrot, Daniel Kirchhofer, Mark S. Dennis, Lydia Santell, Robert A. Lazarus, Jennifer Stamos, and Mark H. Ultsch. The factor VII zymogen structure reveals reregistration of β strands during activation. *Structure*, 9(7):627–636, 2001. ISSN 0969-2126. doi: 10.1016/S0969-2126(01)00624-4. URL `https://www.sciencedirect.com/science/article/pii/S0969212601006244`.

Gracie Wee Ling Eng and David M Virshup. Site-specific phosphorylation of casein kinase 1 δ (CK1δ) regulates its activity towards the circadian regulator PER2. *PLoS One*, 12 (5):e0177834, 2017.

Denis J. Evans and Brad Lee Holian. The Nose-Hoover thermostat. *Journal of Chemical Physics*, 83(8):4069–4074, 1985. ISSN 00219606. doi: 10.1063/1.449071. URL `http://aip.scitation.org/doi/10.1063/1.449071`.

Jin-Yuan Fan, Fabian Preuss, Michael J Muskus, Edward S Bjes, and Jeffrey L Price. Drosophila and vertebrate casein kinase Iδ exhibits evolutionary conservation of circadian function. *Genetics*, 181(1):139–152, 2009.

Federico Forneris, Daniel Ricklin, Jin Wu, Apostolia Tzekou, Rachel S Wallace, John D Lambris, and Piet Gros. Structures of C3b in complex with factors B and D give insight into complement convertase formation. *Science (New York, N.Y.)*, 330(6012):1816–20, 2010. ISSN 1095-9203. doi: 10.1126/science.1195821. URL `http://www.ncbi.nlm.nih.gov/pubmed/21205667`.

John H. Fountain and Sarah L. Lappin. *Physiology, Renin Angiotensin System*. Stat-Pearls Publishing, Treasure Island (FL), 2021. URL `http://europepmc.org/books/NBK470410`.

Brendan T Freitas, Ian A Durie, Jackelyn Murray, Jaron E Longo, Holden C Miller, David Crich, Robert Jeff Hogan, Ralph A Tripp, and Scott D Pegan. Characterization and noncovalent inhibition of the deubiquitinase and deISGylase activity of SARS-CoV-2 papain-like protease. *ACS Infectious Diseases*, 6(8):2099–2109, 2020.

Daan Frenkel and Berend Smit. *Understanding Molecular Simulation*. Academic Press, 2002.

Natalia Friedland, Heng-Ling Liou, Peter Lobel, and Ann M Stock. Structure of a cholesterol-binding protein deficient in Niemann–Pick type C2 disease. *Proceedings of the National Academy of Sciences*, 100(5):2512–2517, 2003.

Matthew Frieman, Kiira Ratia, Robert E Johnston, Andrew D Mesecar, and Ralph S Baric. Severe acute respiratory syndrome coronavirus papain-like protease ubiquitin-like domain and catalytic domain regulate antagonism of IRF3 and NF-κB signaling. *Journal of Virology*, 83(13):6689–6705, 2009.

M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian~09 Revision E.01, 2009. Gaussian Inc. Wallingford CT.

Ziyang Fu, Bin Huang, Jinle Tang, Shuyan Liu, Ming Liu, Yuxin Ye, Zhihong Liu, Yuxian Xiong, Wenning Zhu, Dan Cao, et al. The complex structure of GRL0617 and SARS-

CoV-2 PLpro reveals a hot spot for antiviral drug discovery. *Nature Communications*, 12(1):1–12, 2021.

Luke J Fulcher and Gopal P Sapkota. Functions and regulation of the serine/threonine protein kinase CK1 family: moving beyond promiscuity. *Biochemical Journal*, 477(23): 4603–4621, 2020.

Xiaopan Gao, Bo Qin, Pu Chen, Kaixiang Zhu, Pengjiao Hou, Justyna Aleksandra Wojdyla, Meitian Wang, and Sheng Cui. Crystal structure of SARS-CoV-2 papain-like protease. *Acta Pharmaceutica Sinica B*, 11(1):237–245, 2021.

Jakob Gebel. CK1 gamma. unpublished, 2022.

Jakob Gebel, Marcel Tuppi, Apirat Chaikuad, Katharina Hötte, Martin Schröder, Laura Schulz, Frank Löhr, Niklas Gutfreund, Franziska Finke, Erik Henrich, et al. p63 uses a switch-like mechanism to set the threshold for induction of apoptosis. *Nature Chemical Biology*, 16(10):1078–1086, 2020.

Charles J Geyer et al. Computing science and statistics: Proceedings of the 23rd Symposium on the Interface. *American Statistical Association, New York*, 156, 1991.

Kimberly Fish Gietzen and David M Virshup. Identification of inhibitory autophosphorylation sites in casein kinase I ε. *Journal of Biological Chemistry*, 274(45):32063–32070, 1999.

Peter Goettig, Hans Brandstetter, and Viktor Magdolen. Surface loops of trypsin-like serine proteases as determinants of function. *Biochimie*, 166:52–76, 2019.

Rajeev Goswami, Subhendu Mukherjee, Gerd Wohlfahrt, Chakshusmathi Ghadiyaram, Jwala Nagaraj, Beeram Ravi Chandra, Ramesh K. Sistla, Leena K. Satyam, Dodheri S. Samiulla, Anu Moilanen, Hosahalli S. Subramanya, and Murali Ramachandra. Discovery of Pyridyl Bis(oxy)dibenzimidamide Derivatives as Selective Matriptase Inhibitors. *ACS Medicinal Chemistry Letters*, 4(12):1152–1157, 2013. doi: 10.1021/ml400213v. URL https://doi.org/10.1021/ml400213v.

Alan H Gradman and Rishi Kad. Renin inhibition in hypertension. *Journal of the American College of Cardiology*, 51(5):519–528, 2008.

Paul R Graves and Peter J Roach. Role of COOH-terminal Phosphorylation in the Regulation of Casein Kinase Iδ (∗). *Journal of Biological Chemistry*, 270(37):21689–21694, 1995.

Yoshimi Endo Greer and Jeffrey S Rubin. Casein kinase 1 delta functions at the centrosome to mediate Wnt-3a–dependent neurite outgrowth. *Journal of Cell Biology*, 192(6): 993–1004, 2011.

Yoshimi Endo Greer, Bo Gao, Yingzi Yang, Andre Nussenzweig, and Jeffrey S Rubin. Lack of casein kinase 1 delta promotes genomic instability-the accumulation of dna damage and down-regulation of checkpoint kinase 1. *PloS One*, 12(1):e0170903, 2017.

Jeanne A Hardy and James A Wells. Searching for new allosteric sites in enzymes. *Current Opinion in Structural Biology*, 14(6):706–715, 2004.

Berk Hess. P-LINCS: A parallel linear constraint solver for molecular simulation. *Journal of Chemical Theory and Computation*, 4(1):116–122, 2008.

Berk Hess, Henk Bekker, Herman JC Berendsen, and Johannes GEM Fraaije. LINCS: a linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.

Merl F Hoekstra, Namrita Dhillon, Gilles Carmel, AJ DeMaggio, RA Lindberg, T Hunter, and J Kuret. Budding and fission yeast casein kinase I isoforms have dual-specificity protein kinase activity. *Molecular Biology of the Cell*, 5(8):877–886, 1994.

P Hof, I Mayr, R Huber, E Korzus, J Potempa, J Travis, JC Powers, and W Bode. The 1.8 A crystal structure of human cathepsin G in complex with Suc-Val-Pro-PheP-(OPh) 2: a Janus-faced proteinase with two opposite specificities. *The EMBO Journal*, 15(20): 5481–5491, 1996.

Nadine Homeyer, Anselm HC Horn, Harald Lanig, and Heinrich Sticht. AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *Journal of Molecular Modeling*, 12(3):281–289, 2006.

Lin Hong and Jordan Tang. Flap position of free memapsin 2 (β-secretase), a model for flap opening in aspartic protease catalysis. *Biochemistry*, 43(16):4689–4695, 2004.

William G Hoover. Canonical dynamics: equilibrium phase-space distributions. *Physical Review A*, 31(3):1695, 1985.

James R Horn and Brian K Shoichet. Allosteric inhibition through core disruption. *Journal of Molecular Biology*, 336(5):1283–1291, 2004.

Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins: Structure, Function and Genetics*, 65(3):712–725, 2006a. ISSN 08873585. doi: 10.1002/prot.21123. URL `https://onlinelibrary.wiley.com/doi/full/10.1002/prot.21123`.

Viktor Hornak, Asim Okur, Robert C Rizzo, and Carlos Simmerling. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proceedings of the National Academy of Sciences*, 103(4):915–920, 2006b.

Chaolin Huang, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang, Guohui Fan, Jiuyang Xu, Xiaoying Gu, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, 395(10223):497–506, 2020.

Susanna Hug. Classical molecular dynamics in a nutshell. *Biomolecular Simulations*, pages 127–152, 2013.

Lijian Hui, Latifa Bakiri, Ewa Stepniak, and Erwin F Wagner. p38α: a suppressor of cell proliferation and tumorigenesis. *Cell Cycle*, 6(20):2429–2433, 2007.

William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 1996. ISSN 02637855. doi: 10.1016/0263-7855(96)00018-5.

Tamer M Ibrahim, Muhammad I Ismail, Matthias R Bauer, Adnan A Bekhit, and Frank M Boeckler. Supporting SARS-CoV-2 papain-like protease drug discovery: In silico methods and benchmarking. *Frontiers in Chemistry*, page 996, 2020.

Rodney E Infante, Michael L Wang, Arun Radhakrishnan, Hyock Joo Kwon, Michael S Brown, and Joseph L Goldstein. NPC2 facilitates bidirectional transfer of cholesterol between NPC1 and lipid bilayers, a step in cholesterol egress from lysosomes. *Proceedings of the National Academy of Sciences*, 105(40):15287–15292, 2008.

Alyssa E Johnson, Jun-Song Chen, and Kathleen L Gould. CK1 is required for a mitotic checkpoint that delays cytokinesis. *Current Biology*, 23(19):1920–1926, 2013.

William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 1983. ISSN 00219606. doi: 10.1063/1.445869.

Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Research on Biomolecules*, 22(12):2577–2637, 1983.

Wilson Karubiu, Soumendranath Bhakat, Lara McGillewie, and Mahmoud ES Soliman. Flap dynamics of plasmepsin proteases: insight into proposed parameters and molecular dynamics. *Molecular BioSystems*, 11(4):1061–1066, 2015.

Gergely Katona, Gunnar I Berglund, Janos Hajdu, László Gráf, and László Szilágyi. Crystal structure reveals basis for the inhibitor resistance of human brain trypsin. *Journal of Molecular Biology*, 315(5):1209–1218, 2002.

Bradley A Katz, Richard Mackman, Christine Luong, Kesavan Radika, Arnold Martelli, Paul A Sprengeler, Jing Wang, Hedy Chan, and Lance Wong. Structural basis for selectivity of a small molecule, S1-binding, submicromolar inhibitor of urokinase-type plasminogen activator. *Chemistry & Biology*, 7(4):299–312, 2000.

Theresa Klemm, Gregor Ebert, Dale J Calleja, Cody C Allison, Lachlan W Richardson, Jonathan P Bernardini, Bernadine GC Lu, Nathan W Kuchel, Christoph Grohmann, Yuri Shibata, et al. Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2. *The EMBO journal*, 39(18):e106275, 2020.

Otmar Klingler, Hans Matter, Manfred Schudok, Monica Donghi, Joerg Czech, Martin Lorenz, Hans Peter Nestler, Hauke Szillat, and Herman Schreuder. Structure-based design of amidinophenylurea-derivatives for factor VIIa inhibition. *Bioorganic & Medicinal Chemistry Letters*, 14(14):3715–3720, 2004. ISSN 0960-894X. doi:

10.1016/j.bmcl.2004.05.006. URL https://doi.org/10.1016/j.bmcl.2004.05.006. CAplus AN 2004:498571; MEDLINE PMID: 15203149 (Journal; Article).

Uwe Knippschild, Andreas Gocht, Sonja Wolff, Nadine Huber, Jürgen Löhler, and Martin Stöter. The casein kinase 1 family: participation in multiple cellular processes in eukaryotes. *Cellular Signalling*, 17(6):675–689, 2005.

Uwe Knippschild, Marc Krüger, Julia Richter, Pengfei Xu, Balbina García-Reyes, Christian Peifer, Jakob Halekotte, Vasiliy Bakulev, and Joachim Bischof. The CK1 family: contribution to cellular stress response and its role in carcinogenesis. *Frontiers in Oncology*, 4: 96, 2014.

Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.

Pancham Lal Gupta and Heather A Carlson. Cosolvent Simulations with Fragment-Bound Proteins Identify Hot Spots to Direct Lead Growth. *Journal of Chemical Theory and Computation*, 2022.

Doriano Lamba, Margit Bauer, Robert Huber, Stephan Fischer, Rainer Rudolph, Ulrich Kohnert, and Wolfram Bode. The 2.3 Å crystal structure of the catalytic domain of recombinant two-chain human tissue-type plasminogen activator. *Journal of Molecular Biology*, 258(1):117–135, 1996.

Benoist Laurent, Matthieu Chavent, Tristan Cragnolini, Anna Caroline E Dahl, Samuela Pasquali, Philippe Derreumaux, Mark SP Sansom, and Marc Baaden. Epock: rapid analysis of protein pocket dynamics. *Bioinformatics*, 31(9):1478–1480, 2015.

Guyan Liang, Suzanne Aldous, Gregory Merriman, Julian Levell, James Pribish, Jennifer Cairns, Xin Chen, Sebastien Maignan, Magali Mathieu, Joseph Tsay, et al. Structure-based library design and the discovery of a potent and selective mast cell β-tryptase inhibitor as an oral therapeutic agent. *Bioorganic & Medicinal Chemistry Letters*, 22(2): 1049–1054, 2012.

Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L. Klepeis, Ron O. Dror, and David E. Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function and Bioinformatics*, 78(8):1950–1958, 2010. ISSN 08873585. doi: 10.1002/prot.22711. URL https://onlinelibrary.wiley.com/doi/full/10.1002/prot.22711.

Ho Yin Lo, Peter A Nemoto, Jin Mi Kim, Ming-Hong Hao, Kevin C Qian, Neil A Farrow, Daniel R Albaugh, Danielle M Fowler, Richard D Schneiderman, E Michael August, et al. Benzimidazolone as potent chymase inhibitor: Modulation of reactive metabolite formation in the hydrophobic (P1) region. *Bioorganic & Medicinal Chemistry Letters*, 21(15):4533–4539, 2011.

Edwige Lorthiois, Karen Anderson, Anna Vulpetti, Olivier Rogel, Frederic Cumin, Nils Ostermann, Stefan Steinbacher, Aengus Mac Sweeney, Omar Delgado, Sha-Mei Liao, et al.

Discovery of highly potent and selective small-molecule reversible factor D inhibitors demonstrating alternative complement pathway inhibition in vivo. *Journal of Medicinal Chemistry*, 60(13):5717–5735, 2017.

Shaoyong Lu, Mingfei Ji, Duan Ni, and Jian Zhang. Discovery of hidden allosteric sites as novel targets for allosteric drug design. *Drug Discovery Today*, 23(2):359–365, 2018.

Chunlong Ma, Michael Dominic Sacco, Zilei Xia, George Lambrinidis, Julia Alma Townsend, Yanmei Hu, Xiangzhi Meng, Tommy Szeto, Mandy Ba, Xiujun Zhang, et al. Discovery of SARS-CoV-2 papain-like protease inhibitors through a combination of high-throughput screening and a FlipGFP-based reporter assay. *ACS Central Science*, 7(7):1245–1260, 2021.

Jesper J. Madsen and Ole H. Olsen. Conformational Plasticity-Rigidity Axis of the Coagulation Factor VII Zymogen Elucidated by Atomistic Simulations of the N-Terminally Truncated Factor VIIa Protease Domain. *Biomolecules*, 11(4):549, 2021. ISSN 2218-273X. doi: 10.3390/biom11040549. URL https://www.mdpi.com/2218-273X/11/4/549.

Shavkat Mamatkulov and Nadine Schwierz. Force fields for monovalent and divalent metal cations in TIP3P water based on thermodynamic and kinetic properties. *The Journal of Chemical Physics*, 148(7):074504, 2018.

Kristin L Meagher, Luke T Redman, and Heather A Carlson. Development of polyphosphate parameters for use with the AMBER force field. *Journal of Computational Chemistry*, 24(9):1016–1025, 2003.

Massimiliano Meli and Giorgio Colombo. A Hamiltonian replica exchange molecular dynamics (MD) method for the study of folding, based on the analysis of the stabilization determinants of proteins. *International Journal of Molecular Sciences*, 14(6):12157–12169, 2013.

Qing-Jun Meng, Larisa Logunova, Elizabeth S Maywood, Monica Gallego, Jake Lebiecki, Timothy M Brown, Martin Sládek, Andrei S Semikhodskii, Nicholas RJ Glossop, Hugh D Piggins, et al. Setting clock speed in mammals: the CK1ϵ tau mutation in mice accelerates circadian pacemakers by selectively destabilizing PERIOD proteins. *Neuron*, 58(1):78–88, 2008.

Shuichi Miyamoto and Peter A Kollman. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8):952–962, 1992.

Debra Mlsna, Arthur F Monzingo, Betsy J Katzin, Stephen Ernst, and Jon D Robertus. Structure of recombinant ricin A chain at 2.3 Å. *Protein Science*, 2(3):429–435, 1993.

Luca Monticelli and D Peter Tieleman. Force fields for classical molecular dynamics. *Biomolecular Simulations*, pages 197–213, 2013.

Herbert Nar, Margit Bauer, Angela Schmid, Jean-Marie Stassen, Wolfgang Wienen, Henning WM Priepke, Iris K Kauffmann, Uwe J Ries, and Norbert H Hauel. Structural basis for inhibition promiscuity of dual specific thrombin and factor Xa blood coagulation inhibitors. *Structure*, 9(1):29–37, 2001.

Uddin Mohammad NAsIR, Fumiaki Suzuki, Takao Nagai, Tsutomu NAKAGAwA, and Yukio NAKAMURA. Tyrosine-83 of human renin contributes to biphasic pH dependence of the renin-angiotensinogen reaction. *Bioscience, Biotechnology, and Biochemistry*, 63 (6):1143–1145, 1999.

David Neidhart, Yaoming Wei, Constance Cassidy, Jing Lin, W Wallace Cleland, and Perry A Frey. Correlation of low-barrier hydrogen bonding and oxyanion binding in transition state analogue complexes of chymotrypsin. *Biochemistry*, 40(8):2439–2447, 2001.

Taavi K Neklesa, James D Winkler, and Craig M Crews. Targeted protein degradation by PROTACs. *Pharmacology & Therapeutics*, 174:138–144, 2017.

Weiling Niu, Zhiwei Chen, Prafull S. Gandhi, Austin D. Vogt, Nicola Pozzi, Leslie A. Pelc, Fatima Zapata, and Enrico Di Cera. Crystallographic and kinetic evidence of allostery in a trypsin-like protease. *Biochemistry*, 50(29):6301–6307, 2011. ISSN 0006-2960. doi: 10.1021/bi200878c. URL https://pubs.acs.org/doi/10.1021/bi200878c.

Shuichi Nosé. A unified formulation of the constant temperature molecular dynamics methods. *Journal of Chemical Physics*, 81(1):511–519, 1984. ISSN 00219606. doi: 10.1063/1.447334. URL http://aip.scitation.org/doi/10.1063/1.447334.

Shuichi Nosé. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81(1):511–519, 1984.

Hugh Nymeyer. How efficient is replica exchange molecular dynamics? An analytic approach. *Journal of Chemical Theory and Computation*, 4(4):626–636, 2008.

Vladimiras Oleinikovas, Giorgio Saladino, Benjamin P Cossins, and Francesco L Gervasio. Understanding cryptic pocket formation in protein targets by enhanced sampling simulations. *Journal of the American Chemical Society*, 138(43):14257–14263, 2016.

Sjur Olsnes and Jurij V Kozlov. Ricin. *Toxicon*, 39(11):1723–1728, 2001.

J Osipiuk, R Jedrzejczak, C Tesar, M Endres, L Stols, G Babnigg, Y Kim, K Michalska, and A Joachimiak. The crystal structure of papain-like protease of SARS CoV-2. *RSCB PDB*, 10, 2020.

Jerzy Osipiuk, Saara-Anne Azizi, Steve Dvorkin, Michael Endres, Robert Jedrzejczak, Krysten A Jones, Soowon Kang, Rahul S Kathayat, Youngchang Kim, Vladislav G Lisnyak, et al. Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nature Communications*, 12(1):1–9, 2021.

Matthew E Papaconstantinou, Christopher J Carrell, Agustin O Pineda, Kevin M Bobofchak, F Scott Mathews, Christodoulos S Flordellis, Michael E Maragoudakis, Nikos E Tsopanoglou, and Enrico Di Cera. Thrombin functions through its RGD sequence in a non-canonical conformation. *Journal of Biological Chemistry*, 280(33): 29393–29396, 2005.

Courtney L Parke, Edward J Wojcik, Sunyoung Kim, and David K Worthylake. ATP hydrolysis in Eg5 kinesin involves a catalytic two-water mechanism. *Journal of Biological Chemistry*, 285(8):5859–5867, 2010.

M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, 1981. ISSN 00218979. doi: 10.1063/1.328693. URL `http://aip.scitation.org/doi/10.1063/1.328693`.

John M Peters, Renée M McKay, James P McKay, and Jonathan M Graff. Casein kinase I transduces Wnt signals. *Nature*, 401(6751):345–350, 1999.

Jonathan M Philpott, Rajesh Narasimamurthy, Clarisse G Ricci, Alfred M Freeberg, Sabrina R Hunt, Lauren E Yee, Rebecca S Pelofsky, Sarvind Tripathi, David M Virshup, and Carrie L Partch. Casein kinase 1 dynamics underlie substrate selectivity and the PER2 circadian phosphoswitch. *Elife*, 9:e52343, 2020.

Stefano Piana, Alexander G. Donchev, Paul Robustelli, and David E. Shaw. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *Journal of Physical Chemistry B*, 119(16):5113–5123, 2015. ISSN 15205207. doi: 10.1021/jp508971m.

Fabio Pietrucci, Fabrizio Marinelli, Paolo Carloni, and Alessandro Laio. Substrate binding mechanism of HIV-1 protease from explicit-solvent atomistic simulations. *Journal of the American Chemical Society*, 131(33):11811–11818, 2009.

Agustin O. Pineda, Savvas N. Savvides, Gabriel Waksman, and Enrico Di Cera. Crystal structure of the anticoagulant slow form of thrombin. *Journal of Biological Chemistry*, 277(43):40177–40180, 2002. ISSN 0021-9258. doi: 10.1074/JBC.C200465200. URL `https://www.sciencedirect.com/science/article/pii/S0021925819721097`.

Agustin O. Pineda, Zhi-Wei Chen, Alaji Bah, Laura C. Garvey, F. Scott Mathews, and Enrico Di Cera. Crystal structure of thrombin in a self-inhibited conformation. *Journal of Biological Chemistry*, 281(43):32922–32928, 2006. ISSN 0021-9258. doi: 10.1074/JBC.M605530200. URL `https://www.sciencedirect.com/science/article/pii/S0021925820868677`.

Nuria Plattner and Frank Noé. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nature Communications*, 6(1):7653, 2015. ISSN 2041-1723. doi: 10.1038/ncomms8653. URL `http://www.nature.com/articles/ncomms8653`.

Jay W Ponder and David A Case. Force fields for protein simulations. *Advances in Protein Chemistry*, 66:27–85, 2003.

Kiira Ratia, Scott Pegan, Jun Takayama, Katrina Sleeman, Melissa Coughlin, Surendranath Baliji, Rima Chaudhuri, Wentao Fu, Bellur S Prabhakar, Michael E Johnson, et al. A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. *Proceedings of the National Academy of Sciences*, 105(42):16119–16124, 2008.

Neil D Rawlings and Alan J Barrett. [2] families of serine peptidases. *Methods in Enzymology*, 244:19–61, 1994.

María V Revuelta, Jan AL van Kan, John Kay, and Arjen Ten Have. Extensive expansion of A1 family aspartic proteinases in fungi revealed by evolutionary analyses of 107 complete eukaryotic proteomes. *Genome Biology and Evolution*, 6(6):1480–1494, 2014.

Paul Robustelli, Stefano Piana, and David E Shaw. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences of the United States of America*, 115(21):E4758–E4766, 2018. ISSN 1091-6490. doi: 10.1073/pnas.1800690115. URL `http://www.ncbi.nlm.nih.gov/pubmed/29735687http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6003505`.

Julien Roche, John M Louis, and Ad Bax. Conformation of inhibitor-free HIV-1 protease derived from NMR spectroscopy in a weakly oriented solution. *ChemBioChem*, 16(2):214–218, 2015.

Edina Rosta and Gerhard Hummer. Error and efficiency of replica exchange molecular dynamics simulations. *The Journal of Chemical Physics*, 131(16):10B615, 2009.

S Kashif Sadiq and Gianni De Fabritiis. Explicit solvent dynamics and energetics of HIV-1 protease flap opening and closing. *Proteins: Structure, Function, and Bioinformatics*, 78(14):2873–2885, 2010.

Andrej Šali and Tom L Blundell. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3):779–815, 1993.

Kenneth E Sawin, Katherine LeGuellec, Michel Philippe, and Timothy J Mitchison. Mitotic spindle organization by a plus-end-directed microtubule motor. *Nature*, 359(6395):540–543, 1992.

Herman Schreuder and Hans Matter. Serine Proteinases from the Blood Coagulation Cascade. In *Structural Biology in Drug Discovery*, pages 395–422. Wiley, 2020. doi: https://doi.org/10.1002/9781118681121.ch17. URL `https://onlinelibrary.wiley.com/doi/10.1002/9781118681121.ch17`.

Hans Martin Senn and Walter Thiel. QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition*, 48(7):1198–1229, 2009.

Yibing Shan, Venkatesh P Mysore, Abba E Leffler, Eric T Kim, Shiori Sagawa, and David E Shaw. How does a small molecule bind at a cryptic binding site? *PLoS Computational Biology*, 18(3):e1009817, 2022.

Zhengnan Shen, Kiira Ratia, Laura Cooper, Deyu Kong, Hyun Lee, Youngjin Kwon, Yangfeng Li, Saad Alqarni, Fei Huang, Oleksii Dubrovskyi, et al. Design of SARS-CoV-2 PLpro Inhibitors for COVID-19 Antiviral Therapy Leveraging Binding Cooperativity. *Journal of Medicinal Chemistry*, 2021.

Donghyuk Shin, Rukmini Mukherjee, Diana Grewe, Denisa Bojkova, Kheewoong Baek, Anshu Bhattacharya, Laura Schulz, Marek Widera, Ahmad Reza Mehdipour, Georg Tascher, et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature*, 587(7835):657–662, 2020.

Yuta Shinohara, Yohei M Koyama, Maki Ukai-Tadenuma, Takatsugu Hirokawa, Masaki Kikuchi, Rikuhiro G Yamada, Hideki Ukai, Hiroshi Fujishima, Takashi Umehara, Kazuki Tainaka, et al. Temperature-sensitive substrate and product binding underlie temperature-compensated phosphorylation in the clock. *Molecular Cell*, 67(5):783–798, 2017.

Rebecca A Shirk and George P Vlasuk. Inhibitors of Factor VIIa/tissue factor. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 27(9):1895–900, 2007. ISSN 1524-4636. doi: https://doi.org/10.1161/ATVBAHA.107.148304. URL `http://www.ncbi.nlm.nih.gov/pubmed/17600222`.

Anna R Sielecki, Koto Hayakawa, Masao Fujinaga, Michael EP Murphy, Marie Fraser, Alastair K Muir, Cynthia T Carilli, John A Lewicki, John D Baxter, and Michael NG James. Structure of recombinant human renin, a target for cardiovascular-active drugs, at 2.5 Å resolution. *Science*, 243(4896):1346–1351, 1989.

Wolfgang Skala, Daniel T Utzschneider, Viktor Magdolen, Mekdes Debela, Shihui Guo, Charles S Craik, Hans Brandstetter, and Peter Goettig. Structure-function analyses of human kallikrein-related peptidase 2 establish the 99-loop as master regulator of activity. *Journal of Biological Chemistry*, 289(49):34267–34283, 2014.

Anders B Sorensen, Jesper J Madsen, L Anders Svensson, Anette A Pedersen, Henrik Østergaard, Michael T Overgaard, Ole H Olsen, and Prafull S Gandhi. Molecular basis of enhanced activity in factor VIIa-trypsin variants conveys insights into tissue factor-mediated allosteric regulation of factor VIIa activity. *Journal of Biological Chemistry*, 291(9):4671–83, 2016. ISSN 1083-351X. doi: 10.1074/jbc.M115.698613. URL `http://www.ncbi.nlm.nih.gov/pubmed/26694616`.

Lukas S Stelzl and Gerhard Hummer. Kinetics from replica exchange molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 13(8):3927–3935, 2017.

Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1-2):141–151, 1999.

Jane E Sullivan, Geoffrey A Holdgate, Douglas Campbell, David Timms, Stefan Gerhardt, Jason Breed, Alexander L Breeze, Alun Bermingham, Richard A Pauptit, Richard A Norman, et al. Prevention of MKK6-dependent activation by binding to p38α MAP kinase. *Biochemistry*, 44(50):16475–16490, 2005.

Laura Tesmer, Hans Matter, Alexander Liesum, Otmar Klingler, Manfred Schudok, Gerhard Hessler, Mehdipour Ahmad R., S. Paul Bajaj, Stefan Güssregen, Gerhard Hummer, and Herman Schreuder. Crystallography and molecular simulations capture S1 pocket collapse in allosteric regulation of factor VIIa and other serine proteases. unpublished, 2022.

Simon Tiede. Investigation of Cryptic Pockets in NPC2, P38α, Ricin, and Eg5. Bachelor's Thesis, Goethe-University Frankfurt, Germany, 2021.

Simon Tiede. Discovery and Investigation of Cryptic Pockets by Molecular Dynamics Simulations. Poster, German Conference on Cheminformatics, 2022.

Gareth A. Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. PLUMED 2: New feathers for an old bird. *Computer Physics Communications*, 185(2):604–613, 2014. ISSN 0010-4655. doi: 10.1016/J.CPC.2013.09.018. URL https://www.sciencedirect.com/science/article/abs/pii/S0010465513003196.

Marcel Tuppi, Sebastian Kehrloesser, Daniel W Coutandin, Valerio Rossi, Laura M Luh, Alexander Strubel, Katharina Hötte, Meike Hoffmeister, Birgit Schäfer, Tiago De Oliveira, et al. Oocyte DNA damage quality control requires consecutive interplay of CHK2 and CK1 to activate p63. *Nature Structural & Molecular Biology*, 25(3):261–269, 2018.

Netanel Tzarum, Yael Eisenberg-Domovich, Joell J Gills, Phillip A Dennis, and Oded Livnah. Lipid molecules induce p38α activation via a novel molecular switch. *Journal of Molecular Biology*, 424(5):339–353, 2012.

Haralambos Tzoupis, Georgios Leonis, Grigorios Megariotis, Claudiu T Supuran, Thomas Mavromoustakos, and Manthos G Papadopoulos. Dual inhibitors for aspartic proteases HIV-1 PR and renin: advancements in AIDS–hypertension–diabetes linkage via molecular dynamics, inhibition assays, and binding free energy calculations. *Journal of Medicinal Chemistry*, 55(12):5784–5796, 2012.

Peter MU Ung, Phani Ghanakota, Sarah E Graham, Katrina W Lexa, and Heather A Carlson. Identifying binding hot spots on protein surfaces by mixed-solvent molecular dynamics: HIV-1 protease as a test case. *Biopolymers*, 105(1):21–34, 2016.

Kanagasabai Vadivel, Herman A Schreuder, Alexander Liesum, Amy E Schmidt, Gunaseelan Goldsmith, and S Paul Bajaj. Sodium-site in serine protease domain of human coagulation factor IX a: evidence from the crystal structure and molecular dynamics simulations study. *Journal of Thrombosis and Haemostasis*, 17(4):574–584, 2019.

Sandor Vajda, Dmitri Beglov, Amanda E Wakefield, Megan Egbert, and Adrian Whitty. Cryptic binding sites on proteins: definition, detection, and druggability. *Current Opinion in Chemical Biology*, 44:1–8, 2018.

Kenno Vanommeslaeghe, Olgun Guvench, et al. Molecular mechanics. *Current Pharmaceutical Design*, 20(20):3281–3292, 2014.

Austin D. Vogt, Alaji Bah, and Enrico Di Cera. Evidence of the E*-E equilibrium from rapid kinetics of Na$^+$ binding to activated protein C and factor Xa. *Journal of Physical Chemistry B*, 114(49):16125–16130, 2010. ISSN 1520-6106. doi: 10.1021/jp105502c. URL https://pubs.acs.org/doi/10.1021/jp105502c.

Travis T Wager, Ramalakshmi Y Chandrasekaran, Jenifer Bradley, David Rubitski, Helen Berke, Scot Mente, Todd Butler, Angela Doran, Cheng Chang, Katherine Fisher, et al. Casein kinase 1δ/ε inhibitor PF-5006739 attenuates opioid drug-seeking behavior. *ACS Chemical Neuroscience*, 5(12):1253–1265, 2014.

Junmei Wang, Piotr Cieplak, and Peter A Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, 21(12):1049–1074, 2000.

Junmei Wang, Wei Wang, Peter A Kollman, and David A Case. Antechamber: an accessory software package for molecular mechanical calculations. *Journal of the American Chemical Society*, 222:U403, 2001.

Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general Amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004. ISSN 01928651. doi: 10.1002/jcc.20035. URL https://onlinelibrary.wiley.com/doi/full/10.1002/jcc.20035.

Tsjerk Wassenaar. rtc2018. https://www.github.com/Tsjerk/gromacs, 2018.

Susan R Weiss and Julian L Leibowitz. Coronavirus pathogenesis. *Advances in Virus Research*, 81:85–164, 2011.

Armin Welker, Christian Kersten, Christin Müller, Ramakanth Madhugiri, Collin Zimmer, Patrick Müller, Robert Zimmermann, Stefan Hammerschmidt, Hannah Maus, John Ziebuhr, et al. Structure-Activity Relationships of Benzamides and Isoindolines Designed as SARS-CoV Protease Inhibitors Effective against SARS-CoV-2. *ChemMedChem*, 16 (2):340–354, 2021.

Edward J Wojcik, Rebecca S Buckley, Jessica Richard, Liqiong Liu, Thomas M Huckaba, and Sunyoung Kim. Kinesin-5: cross-bridging mechanism to targeted clinical therapy. *Gene*, 531(2):133–149, 2013.

Di Wu, Tiantian Wu, Qun Liu, and Zhicong Yang. The SARS-CoV-2 outbreak: what we know. *International Journal of Infectious Diseases*, 94:44–48, 2020.

Li Xing, Huey S Shieh, Shaun R Selness, Rajesh V Devraj, John K Walker, Balekudru Devadas, Heidi R Hope, Robert P Compton, John F Schindler, Jeffrey L Hirsch, et al. Structural bioinformatics-based prediction of exceptional selectivity of p38 MAP kinase inhibitor PH-797804. *Biochemistry*, 48(27):6402–6411, 2009.
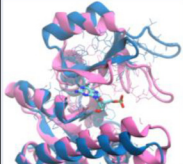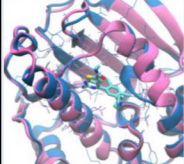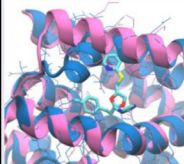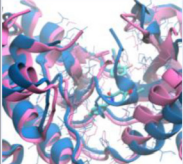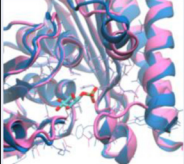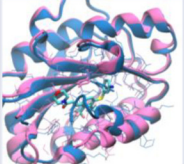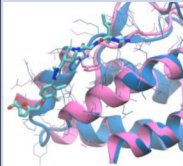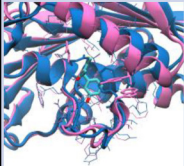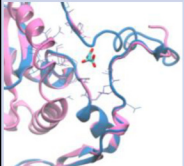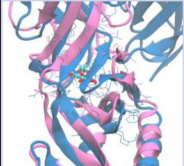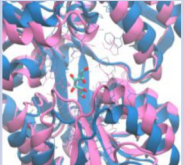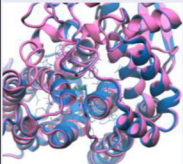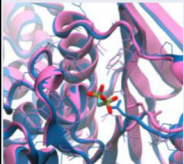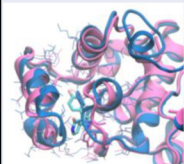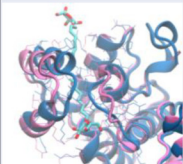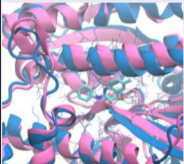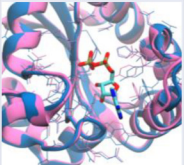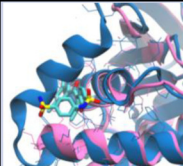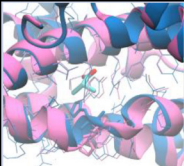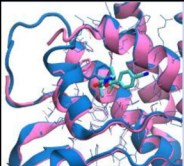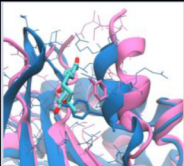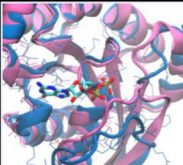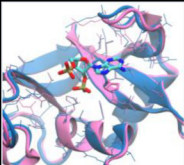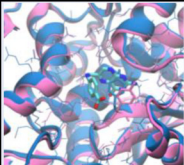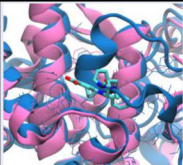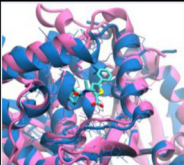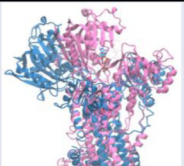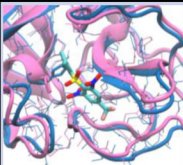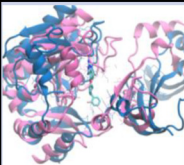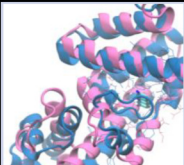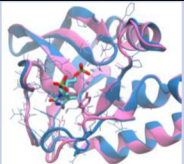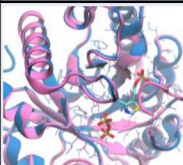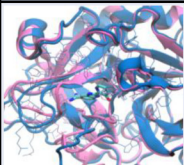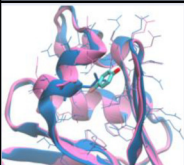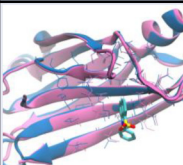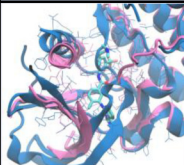
Rui-Ming Xu, Gilles Carmel, Robert M Sweet, Jeff Kuret, and XIAODONG Cheng. Crystal structure of casein kinase-1, a phosphate-directed protein kinase. *The EMBO Journal*, 14(5):1015–1023, 1995.
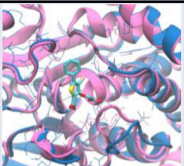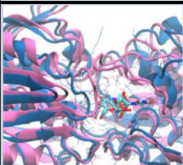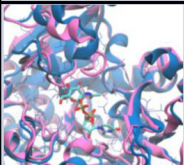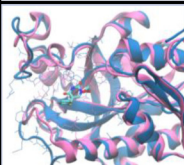
Sujuan Xu, Brian Benoff, Heng-Ling Liou, Peter Lobel, and Ann M Stock. Structural basis of sterol binding by NPC2, a lysosomal protein deficient in Niemann-Pick type C2 disease. *Journal of Biological Chemistry*, 282(32):23525–23531, 2007.

Yechun Xu, Min jun Li, Harry Greenblatt, Wuyan Chen, Aviv Paz, Orly Dym, Yoav Peleg, Tiantian Chen, Xu Shen, Jianhua He, et al. Flexibility of the flap in the active site of BACE1 as revealed by crystal structures and molecular dynamics simulations. *Acta Crystallographica Section D: Biological Crystallography*, 68(1):13–25, 2012.

Xinjian Yan, Thomas Hollis, Maria Svinth, Philip Day, Arthur F Monzingo, George WA Milne, and Jon D Robertus. Structure-based identification of a ricin inhibitor. *Journal of Molecular Biology*, 266(5):1043–1049, 1997.

Youwei Yan, Vinod Sardana, Bei Xu, Carl Homnick, Wasyl Halczenko, Carolyn A Buser, Michael Schaber, George D Hartman, Hans E Huber, and Lawrence C Kuo. Inhibition of a mitotic motor protein: where, how, and conformational consequences. *Journal of Molecular Biology*, 335(2):547–554, 2004.

Yu Yang, Tingting Xu, Yunfei Zhang, and Ximing Qin. Molecular basis for the regulation of the circadian clock kinases CK1δ and CK1ε. *Cellular Signalling*, 31:58–65, 2017.

Yuqi Yu, Jinan Wang, Zhaoqiang Chen, Guimin Wang, Qiang Shao, Jiye Shi, and Weiliang Zhu. Structural insights into HIV-1 protease flap opening processes and key intermediates. *RSC advances*, 7(71):45121–45128, 2017.

Kit-San Yuen, Zi-Wei Ye, Sin-Yee Fung, Chi-Ping Chan, and Dong-Yan Jin. SARS-CoV-2 and COVID-19: The most important research questions. *Cell & Bioscience*, 10(1):1–5, 2020.

Ali M Zaki, Sander Van Boheemen, Theo M Bestebroer, Albert DME Osterhaus, and Ron AM Fouchier. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New England Journal of Medicine*, 367(19):1814–1820, 2012.

Katrin Groebke Zbinden, Ulrike Obst-Sander, Kurt Hilpert, Holger Kühne, David W. Banner, Hans-Joachim Böhm, Martin Stahl, Jean Ackermann, Leo Alig, Lutz Weber, Hans Peter Wessel, Markus A. Riederer, Thomas B. Tschopp, and Thierry Lavé. Selective and orally bioavailable phenylglycine tissue factor/factor VIIa inhibitors. *Bioorganic & Medicinal Chemistry Letters*, 15(23):5344–5352, 2005. ISSN 0960-894X. doi: https://doi.org/10.1016/J.BMCL.2005.04.079. URL https://www.sciencedirect.com/science/article/abs/pii/S0960894X05011169.

Aiwu Zhou, Robin W Carrell, Michael P Murphy, Zhenquan Wei, Yahui Yan, Peter LD Stanley, Penelope E Stein, Fiona Broughton Pipkin, and Randy J Read. A redox switch in angiotensinogen modulates angiotensin release. *Nature*, 468(7320):108–111, 2010.

Na Zhu, Dingyu Zhang, Wenling Wang, Xingwang Li, Bo Yang, Jingdong Song, Xiang Zhao, Baoying Huang, Weifeng Shi, Roujian Lu, et al. A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*, 2020.

Daniel M Zuckerman and Edward Lyman. A second look at canonical sampling of biomolecules using replica exchange simulation. *Journal of chemical theory and computation*, 2(4):1200–1202, 2006.

Appendix

GUANYLATE KINASE	HEAT SHOCK PROTEIN HSP 90-ALPHA	HEME OXYGENASE 1	HEPATOCYTE GROWTH FACTOR RECEPTOR	INOSITOL MONOPHOSPHATASE	INTEGRIN ALPHA-L
INTERLEUKIN 2	KINESIN-LIKE PROTEIN KIF11 (Eg5)	KYNURENINE ALPHA-AMINOADIPATE_AMINOTRANSFERASE	LYSOZYME	MAJOR ENVELOPE PROTEIN E	MALTODEXTRIN-BINDING PROTEIN
METHANE MONOOXYGENASE COMPONENT A ALPHA CHAIN	MICROCIN C7 ANALOG	MITOGEN-ACTIVATED PROTEIN KINASE	MITOGEN-ACTIVATED PROTEIN KINASE 14 (p38α)	MYOSIN II HEAVY CHAIN	NUCLEOSIDE DIPHOSPHATE KINASE A
PROTEIN-TYROSINE PHOSPHATASE	PYRUVATE DEHYDROGENASE LIPOAMIDE KINASE ISOZYME 1	PYRUVATE DEHYDROGENASE KINASE ISOENZYME 2	PYRUVATE KINASE	RAT ADP-RIBOSYLATION FACTOR-1	RHO GTPASE ACTIVATING PROTEIN
RHO-GTPASE-ACTIVATING PROTEIN 1	RIBONUCLEASE A	RICIN	RNA-DIRECTED RNA POLYMERASE	RNA-DIRECTED RNA POLYMERASE 2nd	SARCOPLASMIC ENDOPLASMIC RETICULUM CALCIUM ATPASE
SARS CORONAVIRUS MAIN PROTEINASE	SCAVENGER MRNA-DECAPPING ENZYME DCPS	SERUM TRANSFERRIN	SHIKIMATE KINASE	TETR FAMILY TRANSCRIPTIONAL REPRESSOR LFRR	THERMONUCLEASE
THIAMIN PHOSPHATE SYNTHASE	THROMBIN	TOLUENE-4-MONOOXYGENASE SYSTEM PROTEIN D	TRANSTHYRETIN	TYROSINE-PROTEIN KINASE TIE-2	TYROSINE-PROTEIN PHOSPHATASE NON-RECEPTOR TYPE 1
UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE	UDP-N-ACETYLGLUCOSAMINE PYROPHOSPHORYLASE	URIDINE PHOSPHORYLASE			

Publications

S. N. Cullati, A. Chaikuad, J. Chen, J. Gebel, **L. Tesmer**, R. Zhubi, J. Navarrete-Perea, R. X. Guillen, S. P. Gygi, G. Hummer, V. Dötsch, S. Knapp, and K. L. Gould, *Kinase domain autophosphorylation rewires the activity and substrate specificity of CK1 enzymes*, Molecular Cell, **2022**, 1097-2765

Contributions:

- Set up and performed molecular dynamics simulations of the casein kinase 1 in the native and phosphorylated form

- Connected simulation results with experimental data, i.e., analyzed the stability of the αG helix responsible for the reduction in activity

- Drafted computational part of the manuscript, prepared figures showing the simulations results, edited and reviewed manuscript text

D. Shin, R. Mukherjee, D. Grewe, D. Bojkova, K. Baek, A. Bhattacharya, **L. Schulz**, M. Widera, A. R. Mehdipour, G. Tascher, Georg, P. P. Geurink, A. Wilhelm, G. J. van der Heden van Noort, H. Ovaa, S. Müller, K.-P. Knobeloch, K. Rajalingam, B. A. Schulman, J. Cinatl, G. Hummer, S. Ciesek, and I. Dikic, *Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity*, Nature, **2020**, 587(7835), 657–662

Contributions:

- Set up and performed molecular dynamics simulations of the papain-like protease in complex with 1) ubiquitin, 2) ISG15, and 3) the inhibitor GRL-0617

- Connected simulation results with experimental data, i.e., analyzed the dissociation mechanism of ubiquitin, which explains the reduction of ubiquitin cleavage

- Drafted computational part of the manuscript, prepared figures showing the simulations results, edited and reviewed manuscript text

J. Gebel, M. Tuppi, A. Chaikuad, K. Hötte, M. Schröder, **L. Schulz**, F. Löhr, N. Gutfreund, F. Finke, E. Henrich, J. Mezhyrova, R. Lehnert, F. Pampaloni, G. Hummer, E. H. K. Stelzer, S. Knapp, and V. Dötsch, *p63 uses a switch-like mechanism to set the threshold for induction of apoptosis*, Nature Chemical Biology, **2020**, 16(10), 1078–1086

Contributions:

- Set up and performed molecular dynamics simulations of the casein kinase 1 δ-TAp63α complex, with and without point mutations

- Connected simulation results with experimental data, i.e., analyzed the enzyme-substrate interactions decelerating the phosphorylation kinetics

- Drafted computational part of the manuscript, prepared figures showing the simulations results, edited and reviewed manuscript text

Publication not covered in this Thesis

A. Bazzone, **L. Tesmer**, D. Kurt, H. R. Kaback, K. Fendler, and M. G. Madej, *Investigation of sugar binding kinetics of the E. coli sugar/H+ symporter XylE using solid-supported membrane-based electrophysiology*, Journal of Biological Chemistry, **2022**, 298(2).

Talks

Cryptic pockets - renin and factor VIIa revisited. German Conference on Cheminformatics, virtual, November 2020.

Sampling of ligand-induced conformational changes in renin and factor VIIa. Molecular Modelling Workshop, Erlangen, Germany, February 2020. *(award for 3rd best talk by a PhD candidate)*

Poster presentations

Laura Tesmer, Ahmad Reza Mehdipour, and Gerhard Hummer. Atomistic insight into substrate activity of SARS-CoV-2 papain-like protease and human casein kinase 1. International Conference on Chemical Structures, Noordwijkerhout, Netherlands, June 2022.

Simon Tiede, **Laura Tesmer**, and Gerhard Hummer. Discovery and investigation of cryptic pockets by molecular dynamics simulations. German Conference on Cheminformatics, Garmisch-Partenkirchen, Germany, May 2022.

Laura Schulz, Ahmad Reza Mehdipour, and Gerhard Hummer. Sampling of ligand-induced conformational changes in renin and factor VIIa. Molecular Modelling Workshop, Erlangen, Germany, February 2020.

Laura Schulz, Ahmad Reza Mehdipour, and Gerhard Hummer. Sampling of ligand-induced protein conformational changes. German Conference on Cheminformatics, Mainz, Germany, November 2019.

Laura Schulz, Ahmad Reza Mehdipour, and Gerhard Hummer. Sampling of ligand-induced protein conformational changes. Workshop on Computer Simulation and Theory of Macromolecules, Hünfeld, Germany, March 2019.

Laura Schulz, Ahmad Reza Mehdipour, and Gerhard Hummer. Study of protein allostery. CECAM Workshop "MolSim", Amsterdam, Netherlands, January 2019.

Teaching

- IMPRS workshop on MD simulations, MPI of Biophysics (Oct 2021)

- "Computational Drug Design", lectures and practical course - together with Dr. F. Blanc and Dr. R. Bhaskara, GU Frankfurt (WS 20/21)

- "Versuch 13", practical course on MD simulations for students of biophysics, GU Frankfurt (WS 20/21)

- "Computational Drug Design", practical course, GU Frankfurt (WS 19/20)

- "Seminar Biophysik", summer school in Riezlern for students of biophysics, GU Frankfurt (SS 19)

- "Introduction to MD simulations", practical course, GU Frankfurt (SS 19)

- "Computational Drug Design", single guest lecture and practical course, GU Frankfurt (WS 18/19)

Supervision of students

- Supervision of Simon Tiede, Bachelor Thesis student "Investigation of Cryptic Pockets in NPC2, p38α, ricin, and Eg5 - using unbiased MD simulations, MD simulations with probes and SWISH simulations" and working student (November 2020 to January 2022)

- Co-supervision of Tim Gemeinhardt, Master Thesis student "Experimental and computational studies on the proton/uracil symporter UraA", supervision together with Dr. Ahmad R. Mehdipour and in collaboration with Prof. Eric Geertsma, Goethe University, Frankfurt (April - August 2019)

Eidesstattliche Versicherung

Erklärung

Ich erkläre hiermit, dass ich mich bisher keiner Doktorprüfung unterzogen habe.

Frankfurt am Main, den

_____

Laura Tesmer

Eidesstattliche Versicherung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel

Simulations of Conformational Changes and Enzyme-Substrate Interaction in Protein Drug Targets

selbstständig angefertigt und mich keiner anderen Hilfsmittel als den darin angegebenen bedient habe. Insbesondere haben schriftliche Entlehnungen nicht stattgefunden, soweit sie nicht in dieser Arbeit ausdrücklich mit entsprechendem Quellenverweis angegeben sind.

Ich versichere, die Grundsätze der guten wissenschaftlichen Praxis beachtet, und nicht die Hilfe einer kommerziellen Promotionsvermittlung in Anspruch genommen zu haben.

Frankfurt am Main, den

_____

Laura Tesmer