All-Atom Modeling of Protein Folding and Aggregation

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der Fakultät für Physik des
Karlsruher Institut für Technologie (KIT)

genehmigte

DISSERTATION

von
Dipl.-Phys. Moritz Wolf
aus Karlsruhe

Tag der mündlichen Prüfung: 22. November 2013
Referent: Prof. Dr. Wolfgang Wenzel
Korreferent: Prof. Dr. G. Ulrich Nienhaus

Deutsche Zusammenfassung

Neben etwa 70% Wasser stellen Proteine den Großteil der Masse einer Zelle dar und bestimmen damit deren Aufbau. Strukturell sind sie die komplexesten bekannten Moleküle, welche perfekt auf ihre jeweiligen Funktionen abgestimmt sind. Ihre Struktur hat sich innerhalb eines evolutionären Prozesses weiter entwickelt. So besitzt ein menschliches Chromosom in etwa die gleiche Anzahl an Genen wie das einer Schaumkresse, jedoch sind die menschlichen Proteine im Durchschnitt weitaus komplexer aufgebaut. Innerhalb der Zelle kontrollieren Proteine den Teilchentransport durch die Zellmembran, leiten Signale an den Zellkern weiter, oder regeln den Metabolismus. Andere Proteine wirken als Toxine, dienen als Antikörper zur Abwehr von Infektionen, oder steuern als Hormone wichtige Körperfunktionen.
Thermodynamisch betrachtet befindet sich die native Proteinkonformation im globalen Minimum der Gibbs Freien Energie. Nachdem ein Protein am Ribosom synthetisiert wurde, nimmt es selbständig, oder mit Hilfe weiterer Proteine (Chaperonen), diese eindeutige dreidimensionale Struktur an, die allein von seiner Primärsequenz abhängt.
Durch Sequenzierung von DNA kann diese Sequenz bestimmt werden, jedoch ist bei derzeit rund 86.000 Einträgen in der Proteindatenbank (PDB) nur ein Bruchteil der dazugehörigen Strukturen bekannt. Dies rührt daher, dass die Struktur eines Proteins nur aufwändig experimentell bestimmt werden kann, in der Regel mittels Röntgenbeugung oder NMR-Messungen. Außerdem können nicht alle Proteine für die Röntgenbeugung kristallisiert werden, andere nur mit einhergehender Strukturveränderung, die der Umgebungswechsel nach sich zieht. Auch gestaltet sich die Analyse der gewonnenen NMR-Daten für Proteinkomplexe mit mehr als 200 Aminosäureresten als sehr schwierig. Für viele Proteine verschließt sich die Struktur auch komplett dem Zugang. Gerade innerhalb einer Membran ist es oft nicht möglich, strukturelle Details eines Proteins zu messen. Allerdings machen Membranproteine rund 25% des Genoms eines Eukaryoten aus und sind für die meisten Membranfunktionen einer Zelle verantwortlich. Obwohl sie etwa die Hälfte aller Ansatzpunkte von Medikamenten darstellen, befinden sich derzeit nur 237 Strukturen mit einer experimentellen Auflösung unter 2.5Å in der PDB.

Die Struktur eines Proteins ist jedoch der Schlüssel zu seiner Funktionalität. Fehlerhafte Strukturen beeinträchtigen diese Funktionalität und können oft in Zusammenhang mit Krankheiten, den Proteopathien, gebracht werden. So werden Proteine mit strukturellen Defekten entweder direkt abgebaut, oder innerhalb der Zelle eingelagert. Im ersten Fall können physiologische Funktionsverluste zu Krankheiten wie der Mukoviszidose führen. Der zweite Fall führt hauptsächlich zu neurodegenerativen Krankheiten wie Parkinson oder Alzheimer, in deren Verlauf sich die abgelagerten Proteine zu plaqueartigen Gebilden mit toxischer Wirkung zusammenschließen.

Computergestützte Methoden können hilfreiche strukturelle Informationen über Proteine liefern und darüber hinaus auch Einblicke in deren Dynamik geben. Jedoch sind diese Simulationen

aufgrund der Komplexität des Problems auf sehr kleine Zeitskalen beschränkt.
Mit Hilfe von speziell für die Molekulardynamik gefertigten Superrechnern ist es nun möglich bis in den Millisekundenbereich vorzudringen und den Faltungsprozess von schnell faltenden Proteinen zu simulieren. Neben der aufwändigen Energieberechnung ist der Femtosekunden-Integrationsschritt der limitierende Faktor in der Molekulardynamik, welcher verhindert größere Zeitskalen zu erreichen. Dieser muss in der Größenordnung des schnellsten Übergangs im System, der atomaren Schwingungen, liegen.
Eine stochastische Herangehensweise ist nicht an diese Oszillationen gebunden und ermöglicht die effizientere Beschreibung eines Proteins in einem thermodynamischen Zustandsensemble. Dies ermöglicht zwar die Beschreibung großer Konformationsübergänge, jedoch geht damit ebenfalls der Verlust der kinetischen Information einher. Trotz diesem erheblichen Vorteil ist die Beschreibung der Proteinfaltung mittels Monte Carlo Methoden bisher nur für kleine Proteine möglich und ebenfalls mit dem Einsatz immenser Rechenkapazitäten verknüpft.

Ziel dieser Arbeit ist es, die Simulation von komplexen Proteinsystemen voranzutreiben, um Einblicke in bisher verborgene Prozesse zu ermöglichen. Dies umfasst zum einen die Entwicklung eines effizienten Kraftfeldes, um mit Hilfe der Molekulardynamik die Assemblierung komplexer Porensysteme innerhalb einer Membran zu simulieren und zum anderen die Entwicklung eines effizienten Monte Carlo basierten Simulationsprogramms, das die Proteinfaltung in physikalischen Kraftfeldern auf konventioneller Hardware ermöglicht. Weiter werden massiv parallele Methoden untersucht, implementiert und weiterentwickelt, damit diese auf Proteinsysteme angewandt werden können. Dies ermöglicht erstmals die reversible Faltung schnell faltender Proteine innerhalb eines Tages zu simulieren.

Assemblierung des Twin Arginine Translocase Porenkomplexes

Der Proteintransport durch Zellmembranen ist eine fundamentale Eigenschaft lebender Zellen, der durch die hohe Anzahl an für Transportproteine codierenden Genen verdeutlicht wird. Das Twin-Arginine-Translocase-System (Tat) unterscheidet sich von anderen Transportsystemen dahingehend, dass es statt unstrukturierter Aminosäurefäden komplett gefaltete Proteine durch Membranen transportiert. Aus diesem Grund muss es weitaus größere Poren innerhalb einer Membran bilden als z.B. das Sec-System[1]. Insbesondere für den photosynthetischen Elektronentransfer, welcher in Bakterien über die Zellmembran hinweg stattfindet, ist dies von großer Bedeutung. Die hierfür benötigten redox-aktiven Proteine erhalten erst durch Binden eines Kofaktors ihre redox-aktiven Eigenschaften. Da dies in der Regel nur durch Falten des Protein ermöglicht wird, kann der resultierende Komplex aufgrund seiner Größe nicht mehr über den Sec-Pfad durch die Zellmembran transportiert werden.
Das Tat System bietet hierfür einen Ausweg. Bestehend aus einer Vielzahl an Einzelproteinen der beiden Typen TatA und TatC, formt es eine Pore passender Größe dynamisch um das zu transportierende Cargoprotein herum und löst sich nach dem Transport, ohne Schänden an der Zelle zu hinterlassen, wieder auf. Ausgelöst wird dieser Prozess durch ein Signalpeptid, welches zum einen an das Cargoprotein bindet und zum anderen anhand eines *Twin Arginine* Musters von einem TatC-Protein detektiert wird. Daraufhin wird die Pore aus mehreren TatA-

Einheiten um das Cargo herum geformt. Je nach Organismus ist ein weiteres TatA-Homolog, das TatB-Protein, an diesem Prozess beteiligt, jedoch in weitaus geringerer Stückzahl. Jedes TatA-Monomer besteht aus einem Transmembransegment (TMS), einer amphipathischen Helix (APH) und einem dicht geladenen, unstrukturierten Rest (DCR).

Innerhalb einer Kollaboration mit der Arbeitsgruppe Ulrich am KIT und der Arbeitsgruppe Ruggerone an der Università di Cagliari, haben wir ein Modell für die treibende Kraft hinter dem Assemblierungsprozess des $TatA_d/C_d$ -Porenkomplex im Bacillus Subtilis Organismus entwickelt. Wir haben festgestellt, dass die porenbildenden $TatA_d$ -Monomere ein intrinsisch komplementäres Ladungsmuster in ihrer Sequenz tragen. Obwohl ein Monomer im Gesamten fast ungeladen ist, weisen seine außerhalb der Membran liegenden Domänen rund 40% geladene Residuen auf, wohingegen, statistisch betrachtet, für Transmembranproteine nur 22.5% zu erwarten wären. Strukturell gesehen bildet das komplementäre Ladungsmuster zwischen DCR und APH eine Art Reißverschluss, stabilisiert durch sieben Salzbrücken. Solche "Charge Zipper" konnten wir ebenfalls bei 75% der Einträge in der Uniprot Datenbank finden und postulieren daher den *Charge Zipper Mechanismus* als generelles Prinzip, verantwortlich für die Selbstassemblierung von Membranproteinen.
Basierend auf den *Charge Zippern* haben wir ein strukturbasiertes Membran/Poren-Potential entwickelt, um erstens deren sterische Plausibilität zu demonstrieren und zweitens die Rolle des Reißverschlussmechanismus als treibende Kraft zur Selbstassemblierung zu stärken. Eine massive Mutationsstudie unserer experimentellen Partner konnte die Bestimmung und Zuweisung der postulierten Salzbrücken untermauern. Die Übertragbarkeit des Kraftfelds wurde durch Simulationen an weiteren interessanten Membranproteinen, dem Stressantwortpeptid TisB, dem antimikrobiellen Peptid Dermcidin und dem pestiviralen E^{rns} Peptid gezeigt.

Basierend auf einem Ladungsmuster aus fünf intra- und zwei intermolekularen Salzbrücken, war es uns möglich, wiederholt den Assemblierungsprozess einer stabilen Pore, bestehend aus zwölf $TatA_d$ -Einheiten, zu simulieren. Die resultierenden Poren stimmen sehr gut mit experimentell gemessenen cryo-EM Daten hinsichtlich Größe und Form überein. Bedingt durch die Heterogenität der TatA-Komplexe, erwies sich deren atomistische Auflösung mittels Röntgenbeugung oder NMR als zu anspruchsvoll, und es konnten bisher nur einzelne TatA-Monomere oder deren Fragmente experimentell bestimmt werden. Unsere Simulationen erlauben daher erstmals genauere Einblicke in die strukturelle Zusammensetzung einer assemblierten TatA-Pore.

Proteinfaltung

Wie bereits beschrieben, ist die Struktur eines Proteins hinreichend durch seine Aminosäuresequenz definiert. Mit Hilfe dieser nativen Struktur lassen sich Aufschlüsse über die Funktion des Proteins geben. Der natürliche Übergang von einer unstrukturierten Konformation in diese funktionelle Struktur wird als Proteinfaltung bezeichnet. Experimentell konnte gezeigt werden, dass Faltung und Entfaltung reversibel durch Manipulation des pH-Werts der Umgebung zu erreichen sind. Während sich ein Protein in der Natur binnen Mikrosekunden bis Sekun-

den faltet, birgt die theoretische Beschreibung dieses Vorgangs zahlreiche Hindernisse. Zwar ist die Darstellung mittels klassischer Mechanik durchaus genügend, jedoch ist die Evaluierung der benötigten physikalischen Kraftfelder mit einem großen Rechenaufwand verknüpft. Auch müssen in der vorherrschenden Simulationstechnik, der Molekulardynamik (MD), schrittweise im Femtosekundenbereich die Newtonschen Bewegungsgleichungen integrativ gelöst werden. Der damit verbundene Rechenaufwand konnte erst vor kurzem von einem eigens für diesen Zweck konzipierten Superrechner aufgebracht werden. Erstmals konnte eine MD-Simulation in den Millisekundenbereich vordringen und die reversible Faltung von schnell faltenden Proteinen wiederholt darstellen.

Unser Ziel ist es, die Proteinfaltung mittels Monte Carlo basierter Simulationstechniken zu ermöglichen. Dafür haben wir im Rahmen dieser Arbeit ein effizientes und flexibles Simulationsprogramm, SIMONA[1], entwickelt, unter dessen Einsatz wir in der Lage waren, ebenfalls mehrfache reversible Faltungs- und Entfaltungsübergänge schnell faltender Proteine zu simulieren. Dabei setzten wir die Parametrisierung des etablierten Molekulardynamik-Kraftfelds AMBER99SB-STAR-ILDN ein. Die wässrige Umgebung berechneten wir implizit unter Verwendung eines effektiven Generalized Born Lösers.

Wir konnten erfolgreich Gleichgewichtssimulationen der schnell faltenden Subdomäne des Villin headpiece durchführen. Dabei beobachteten wir zahlreiche Faltungs- und Entfaltungsübergänge des 37 Residuen langen Proteins, aus denen wir erfolgreich die zu Grunde liegende Freie Energie-Landschaft berechnen konnten. Unsere Messungen der Barrierenhöhen zur Stabilisierung der nativen Konfiguration stimmen gut mit denen aus Molekulardynamiksimulationen und experimentellen Messungen überein. Eine Untersuchung der Thermostabilität einzelner Strukturelemente ergab Aufschlüsse über die Eigenschaften der drei enthaltenen Helixregionen. So konnten wir zeigen, dass eine der Helices im Faltungsübergangszustand unstrukturiert ist. Dies deckte sich ebenfalls mit der Berechnung der ϕ-Werte, aus denen ersichtlich wurde, dass, im Gegensatz zu beiden anderen Helices, diese dem entfalteten Ensemble und nicht der native Konfiguration ähnelt. Im frühen Strukturbildungsprozess konnten wir bereits die Bildung erster helikalen Tertiärkontakte beobachteten, was dem Nukleationsmodell entspricht. Der überwiegend helikale Charakter der simulierten Strukturen zeigte sich auch in der Berechnung der Circulardichroismusspektren durch ausgeprägte Signale in den helikalen Bändern. Durch das breite Spektrum an simulierten Temperaturen, war es uns möglich, die temperaturabhängige Elliptizität zu bestimmen, die in ihrem Verlauf sehr gut mit experimentellen Ergebnissen übereinstimmt. Damit konnten wir zeigen, dass Monte Carlo basierte Methoden durchaus in der Lage sind, eine korrekte thermodynamische Beschreibung komplexer Vorgänge in Proteinen zu liefern. Die für unsere Simulationen benötigte Rechenzeit lag dabei etwa drei Größenordnungen unter der vergleichbarer Molekulardynamik Simulationen. Mit der Veröffentlichung des SIMONA Codes und seiner Freigabe für die akademische Nutzung ermöglichen wir der Community erstmals die effiziente Simulation von Proteinen auf effektiven Zeitskalen im Millisekundenbereich, unter Verwendung herkömmlicher Hardware.

[1] Die Entwicklung des Simulationspakets SIMONA erstreckt sich über die beiden Dissertation von Timo Strunk (C++ Kernel) und Moritz Wolf (Präprozessor, GUI, Multitransformationen). Der Quellcode kann unter `http://int.kit.edu/nanosim/simona` heruntergeladen werden und darf für akademische Zwecke frei verwendet werden.

Multiple Try Metropolis

Trotz der drastischen Reduzierung des Rechenaufwands bei der Benutzung einer stochastischen Simulationsmethode benötigen wir für unsere Villin Simulationen etwa einen Monat um vier Faltungsübergänge beobachten zu können. Durch Parallelisierung der Energieauswertung konnte dies auf eine Woche reduziert werden, jedoch haben schon hoch optimierte MD-Simulationspakete gezeigt, dass diese Art der Optimierung sehr schnell ihr Limit erreicht und nicht für den Einsatz auf Superrechnern geeignet ist. Die Entwicklung einer skalierbaren Monte Carlo-Simulationsmethode für Superrechner würde einen hochgradigen Vorteil gegenüber Molekulardynamiksimulationen erzielen und dessen Anwendungsbreite steigern.
Vielversprechend ist hierbei die Optimierung des Markov-Prozesses selbst. Den geeignetsten Angriffspunkt der streng iterativen Methodik des Metropolis Hastings Algorithmus (MH) bietet der Generatorschritt, in dem ein neuer randomisierter Zustand erzeugt wird. Triviale Änderungen der Schrittweite erlauben zwar ein effizienteres Sampling, doch geht dies mit verringerten Akzeptanzraten einher, was den Raum für Verbesserungen ebenfalls stark einschränkt. Die Multiple-Try-Metropolis-Methode (MTM) sieht eine weitaus effektivere Erweiterung vor und erzeugt anstatt eines einzelnen Zustands mehrere Zustände gleichzeitig. Zwar wird zur Einhaltung des Detailed Balance Kriteriums, vergleichbar mit einer Kontinuitätsgleichung in der Thermodynamik, ein weiterer Generatorschritt benötigt, jedoch sinkt die Autokorrelationszeit zwischen zwei aufeinander folgenden MTM-Schritten derart, dass der Algorithmus stets effizienter ist als ein herkömmlicher MH Ansatz. Die Vorteile dieser Methode liegen auf der Hand, denn es können nun beliebig komplizierte Zustände generiert werden, da mit steigender Anzahl an Vorschlägen auch die Akzeptanzrate steigt. Weiter ist die Methode trivial parallelisierbar und damit bestens für dein Einsatz auf Superrechnern geeignet.

Ein Aspekt dieser Arbeit beschäftigt sich mit der Anwendbarkeit der MTM-Methode auf das Problem der Proteinfaltung. Obwohl dies zu Anfang als trivial erschien, birgt die Übertragung der Methode von mathematischen Problemen auf die Komplexität einer Proteinenergiefunktion gravierende Hürden. Wir konnten diese jedoch erfolgreich ausfindig machen und beseitigen und somit einen Gültigkeitsbereich der Methode aufstellen, innerhalb dessen die Simulation von Proteinen möglich ist. In einer Weiterentwicklung der MTM-Methode konnten wir einen effizienteren MTM-G-Ansatz kreieren, der für die Anwendung auf Proteinsysteme bestens geeignet ist. Somit war es uns möglich Simulationen von Proteinsystemen mit Hilfe der MTM-G-Methode erfolgreich durchzuführen. Hierbei konnten weit mehr Faltungs- und Entfaltungsübergänge pro Simulationsschritt beobachtet werden, als in zuvor durchgeführten MH-Simulationen. Eine Abschätzung des Zeitequivalents durch einen Vergleich mit experimentellen Faltungszeiten ergab eine enorme Effizienz von $17ps$ pro MTM-G Schritt, vier Größenordnungen über dem einer MD-Simulation. Berechnete Freie Energie Landschaften decken sich perfekt mit den vorherigen MH-Untersuchungen. Letztendlich erlaubte uns die enorme Kapazität zur Parallelisierung einer einzelnen MTM-G-Simulation, unter Verwendung eines Superrechners, die Erzeugung von reversiblen Faltungstrajektorien des Villin Headpiece Proteins innerhalb eines Tages. Damit könnte der MTM-G die Grundvoraussetzungen für die Untersuchung längerskaliger Prozesse geschaffen haben.

Contents

1. Introduction

1.1. Overview

Proteins make up the largest fraction of a cell's dry matter. In addition to their structural roles they are also responsible for most of the cellular functionality and, in each of their distinct roles, they are among the best adopted molecules we know. Proteins can regulate the particle transfer across the cell membrane, transport signals to the nucleus or control the cell metabolism. Enzymes can catalyze biochemical reactions, highly specialized proteins operate as toxins or antibodies, while others control as hormones important somatic functions.
Their complex structure has evolved in an evolutionary process and, depending only on its primary sequence, many proteins spontaneously fold into this native configuration. Thermodynamically, the native structure corresponds to the global minimum of the Gibbs free energy and a denatured protein attains its structure on a microsecond to second time scale.
Although the sequence of a protein can be obtained through DNA sequencing with comparatively little effort, measuring its structure proves difficult. By now, the database of known proteins (PDB) contains $86,000$ experimentally resolved structures, a comparatively small part regarding the flood of known sequences. But crystallizing a protein is not always possible and besides, this process may affect the structure of the protein. Also the applications range of NMR is limited by the number of atoms. Especially the PDB lacks of membrane proteins, as it currently contains only 237 of them, measured with an experimental resolution below 2.5Å. However, 25% of a eukaryotic genome code for membrane proteins which are responsible for almost all membrane functions and therefore constitute 50% of current drug targets.

Knowing the structure of a protein is crucial for the understanding of its functionality. Even small structural defects can inactivate a protein. Defect proteins will either be degraded or stored within the cell, which can cause proteopathies. In the former case, the loss of function can result in diseases like Mucoviscidosis, whereas stored proteins could aggregate to plaques causing often neurodegenerative diseases like Parkinson's disease or Alzheimer's disease. In addition, active proteins do not remain in a simple isolated native state, but undergo conformational changes depending on binding partners or charges in the environment. Motor proteins like Myosin for example, are able to hydrolyze ATP to deform and move along Actin filaments, causing large scale muscle contraction.

Computer-based methods can also give insights into the dynamics of proteins. However, due to the complexity of such simulations, they are restricted to small time scales. A modern supercomputer, built to perform a single molecular dynamics (MD) simulation, facilitates the microsecond simulation of proteins in a physical force field. For the first time, it was possible to reversibly fold and unfold several fast-folding proteins multiple times. Besides the sophisticated force field

evaluation, the integration step is the limiting factor of an MD simulation. To conserve the total energy, the integration step has to be smaller than the fastest transition in the system, the atomic vibration. With typical time steps of the order of $1fs$, an MD simulation requires 10^9 energy evaluations to reach the microsecond time scale where the folding transition of the fast-folders occurs. This enormous computational effort often makes the MD unsuitable for the analysis of promising systems.
The stochastic approach of a Monte Carlo (MC) based simulation can potentially overcome these restrictions by describing a thermodynamic conformational ensemble, which is not inherently tied to a particular time scale. MC methods are capable of realizing large conformational changes, however, this entails the loss of the kinetic information. Nevertheless, descriptions from MC simulations often agree well with MD results, because the ergodicity theorem states that the *time average* will converge against the *configuration average* for long simulations. Presently, the MD is the most widely used simulation technique for the description of biomolecular systems or processes. Despite the advantages in phase space exploration, MC simulations still require large computational resources and there are few simulation packages that support simulations to model biomolecular systems. As a consequence, there is a wide variety of MD simulation packages, whereas only few MC frameworks exist.

To overcome this lack of available methods in this thesis an efficient and versatile framework for stochastic simulations of molecular and nanoscale systems was developed. All Monte Carlo based methods presented in this work are implemented in this package, named SIMONA[2, 3] (**SI**mulation of **MO**lecular and **NA**noscale systems[1]). It can be downloaded free of charge for an academic use from `http://int.kit.edu/nanosim/simona`.

1.2. Outline

This work aims to improve the simulation of complex biomolecular systems to acquire insights into processes that cannot be described with established simulation methods on the hardware available to most researchers. The work in this thesis addresses both all-atom simulations for small proteins and the assembly of very large complexes that cannot be tackled even with most advances simulation techniques using biomarker force fields. To address these problems we have developed a fast structure based apporach including an implicit treatment of the environment, in which these aggregation processes take place. In addition, we developed an implicit membrane bilayer model and used it to describe the formation process of large membrane-pore complexes. The development of a Monte Carlo simulation framework allowed us to reversibly fold and unfold small proteins using biomolecular force fields with atomic resolution. In addition to the conventional Metropolis algorithm, we also have developed massively parallel methods to allow an even faster sampling of large scale conformational changes. Using the most sophisticated of these techniques we were able to characterize folding transitions of a fast folding protein within a single day.

[1] The development of the simulation framework SIMONA spans over both dissertations of Timo Strunk (C++ kernel) and Moritz Wolf (preprocessor, GUI, multi transformations).

Twin Arginine Translocase

Protein transport across cell membranes is an important mechanism in cells. The Twin Arginine Translocase (Tat) system represents one particular interesting transport mechanism because instead of transporting unstructured chains of amino acids, the Tat system is capable of transporting fully folded proteins. For this reason it has to selectively form large membrane-spanning pores to accommodate the cargo that must be closed once the transfer process is complete. The significance of such a transport mechanism is important in bacteria that have to transport proteins binding cofactors across the membrane, which requires the protein to fold in the cytosol. Transport of such proteins cannot be accomplished by other mechanisms, such as the Sec pathway.
The Tat system consists of many subunits of the two protein types, TatA and TatC, and can arrange them to dynamically form a pore fitting around a given cargo protein. The pore forming mechanism is initiated after the cargo has bound to a signal peptide, which is recognized by a TatC molecule through a *twin arginine* pattern. Followed by the assembling of a suitable pore consisting of multiple TatA subunits, the cargo protein is transported to the exoplasmic side of the membrane. Afterwards, the pore complex dissolves without harming the membrane. Depending on the organism, other TatA homologs, such as the TatB protein may participate in the transport process.

In cooperation with the group of Anne Ulrich at the KIT and the group of Paolo Ruggerone at the Università di Cagliari, we proposed a novel mechanism for pore formation by TatA[4]. We investigated the Tat complex and noticed an intrinsic complementary charge pattern within the sequences of all TatA homologs. Although its overall charge is nearly zero, a TatA monomer comprises 40% charged residues within extra-membrane regions, whereas only 22.5% would be expected for transport proteins. Within the $\mathrm{TatA_d/C_d}$ complex in the bacillus subtilis organism, the charge pattern manifests itself as ladder of seven pairwise complementary charge pairs between the different structural domains. Based on structural models we postulated that both parts of the molecule attach to each other by successively forming the intramolecular salt bridges in a zipper like manner and adopt a hairpin like structure. By forming the intermolecular salt bridges with adjacent hairpins, a cluster of monomers would then be able to build an amphiphilic palisade with an adequate height to span the membrane. Such palisades could form the translocation pore in the membrane with a hydrophilic interior. In addition to the Tat complex we could find putative *charge zipper motifs* in 75% of the entries of membrane peptides in the Uniprot database. This led to our assumption that *charge zippers* are rather a general mechanism behind the self-assembly of membrane proteins.

To elucidate the important role of the *charge zippers* for the self-assembly of a stable pore system and to demonstrate their structural feasibility, I developed a structure model featuring an implicit membrane-pore potential enriched with explicit attractive constraints for the proposed charge pairs. A thorough mutation study of our experimental partners confirmed the existence of the proposed zipper motif and helped identify the exact salt-bridge pattern with five intramolecular and two intermolecular contacts. Based on this data, we were able to simulate the pore complex. Using a zipper motif comprising five intramolecular and two intermolecular contacts,

we could repeatedly simulate the assembly process of dodecameric $TatA_d$ pores. The resulting pore complexes agreed well with experimental cryo-EM data, regarding their shapes and sizes. Because of the heterogeneity of assembled TatA pores, only low resolution experimental data is available to date. Our simulations thus generate new insights into formation of TatA pore complexes and the Tat translocation mechanism.

The transferability of our approach was demonstrated by simulating the aggregation of other interesting membrane proteins, such as the bacterial stress-response peptide, TisB, the tetrameric assembly of the anionic antimicrobial peptide, Dermcidin, and finally the membrane alignment of the structural glycoprotein of pestiviruses, E^{rns}. Hence, our simulated predictions for the $TatA_d$ pore complex for the first time grant deeper insights into the composition of the Tat translocation mechanism.

Protein Folding

In the folding process, a protein undergoes various conformational changes to transform from an unstructured coil into a stable functional structure. Failure to fold to this native structure, which for many proteins is uniquely defined by the proteins primary amino acid sequence, can result in diseases as the protein cannot accomplish its function. By changing the environmental pH-value, it was shown experimentally that the folding process is reversible in the famous refolding experiments of Anfinsen. Ever then, the details of the folding mechanism of proteins have been investigated and a wide variety of mechanisms have been proposed. The timespan required for folding a protein can vary from the microsecond scale for ultrafast folding proteins up to the second scale for more complex protein structures. In principle, Simulations can aid to understand this process, but only recently a highly specialized supercomputer was able to simulate a reversible folding process of a handful of fast folding proteins using the same biophysics-based approach. In those simulations the millisecond time scale could be reached for the first time using molecular dynamics simulations at the all-atom level. Because the integration step in these simulations is in the femtosecond range, such simulations require an enormous computational cost.
In a thermodynamic view, the native structure of a protein is located at the global minimum of its Gibbs free energy surface. At folding equilibrium the free energies of both states, the enthalpically stabilized native state and the entropically dominated unfolded state, are the same and the system resides in both states with equal probability. This is the optimal temperature for observing multiple transition events in reversible folding simulations.
As an alternative to solving the equations of motion in a molecular dynamics simulation to determine time expectation values, a stochastic Monte Carlo approach can generate the same stationary distribution of the system and thermodynamic observables can be measured without incorporating atomic vibrations. Despite this potential advantage of MC, molecular dynamics is the predominant simulation technique in the biophysical field and only few Monte Carlo based simulation packages are available to explore alternatives.

To facilitate the characterization of the thermodynamic properties of biomolecular systems including reversible folding of proteins with Monte Carlo based methods, we developed the

efficient and versatile simulation framework SIMONA. Using the atomic parametrization of the established molecular dynamics force field AMBER99SB-STAR-ILDN combined with an efficient Generalized Born implicit solvent model to incorporate interactions with the environment, we were able to observe multiple folding and unfolding transitions of the ultrafast folding subdomain of the Villin headpiece.
In our investigations we could characterize the complete thermodynamic landscape of this widely studied fast-folding protein. We could determine the complex free energy landscape and identified a folding intermediate state, separated from the native conformation by a barrier of $1\frac{kcal}{mol}$. In a structural analysis of the relevant conformations of the ensemble, we could characterize the thermal stability of the three helical domains individually. Consistent with experimental observations, a ϕ-value analysis exhibited a high stability of the first two helices and a lower stability for the third one. Based on the ϕ-values analysis, the experimentally inaccessible transition state could be characterized to have non-native regions in the third helix of the protein. We observed that the first tertiary contacts in the unstructured ensemble were formed in a nucleation like process by developing initial helical contacts. Depending on the simulation temperature, the fraction of helical content changes in excellent agreement with experimental observations. By calculating the CD spectrum for all temperature ensembles, we could reproduce the temperature dependent ellipticity of the protein in very good agreement with experimental data. Our contribution to the Monte Carlo based simulation of protein systems allows the efficient characterization of the thermodynamic ensemble of biomolecular systems using off-the-shelf hardware and thus represent a competitive alternative to established MD approaches that can be used only by a single group worldwide at the moment.

Multiple Try Metropolis

Despite the efficiency of our Monte Carlo simulations, simulations to observe reversible folding still require months of real time. 100 million steps in the simulation of the Villin headpiece subdomain required roughly one month simulation time on a single processor. Using an OpenMP approach, we could parallelize the energy evaluation and effectively reduce this computational time to $100Mio$ steps per week including four processors. However, the pursuit of this type of a parallelization of the energy evaluation does not work well for presently available supercomputers. I therefore investigated enhanced Monte Carlo techniques that can exploit these massively parallel architectures to yield an even better acceleration over molecular dynamics simulations.
The Multiple Try Metropolis method (MTM) modifies the proposal step of the Metropolis Hastings method (MH) that successively generates a single random conformation to subsequently accept or reject according to an acceptance criterion. Instead of generation of a single new conformation, MTM generates multiple trial conformations at once. This allows a more complex exploration of the phase space in the vicinity of the current conformation, but seems to incur a large computational overhead. Nevertheless, it was shown for test functions representing mathematical problems that MTM converges faster than the standard MH approach and attains smaller auto-correlation times even when its computational overhead is taken into account.

In this work we investigated the applicability of the MTM method to the simulation of protein

systems. We found that several challenges must be overcome to apply the MTM to these complex high-dimensional problems. However, it was possible to develop adaptations of the method that resulted in a robust and fast simulation technique for proteins. Using this method we were able to simulate the reversible folding process of a protein faster than ever before. We could even observe multiple folding/unfolding events of the Villin headpiece subdomain within a single day.

313, 331, 367, ... ?

"379 ! It's a sequence of happy primes. [...] Any number that reduces to one when you take the sum of the square of its digits in continuing iteration until it yields one is a happy number. Any number that doesn't, isn't. A happy prime is a number that's both happy and prime. [...] Don't they teach recreational mathematics anymore ?"[2]

Doctor Who, 42

[2] Author's note: The largest known happy prime number is: $2^{42643801} - 1$

2. Biomolecular Systems

2.1. Protein Structure

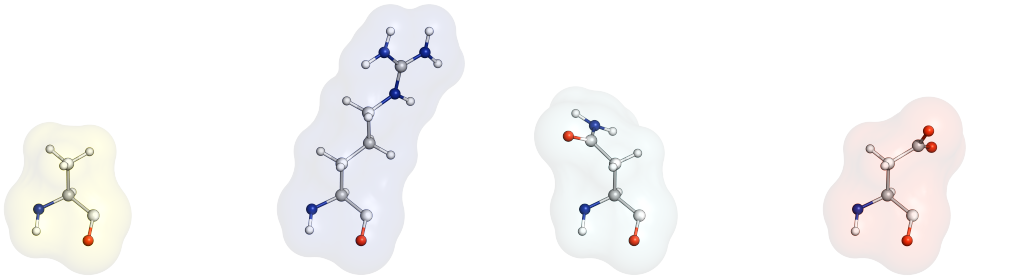
Proteins represent the main constituents of a cells dry matter. In addition to their structural contribution, they are also responsible for the majority of cell functions. This includes the regulation of the cells metabolism and signaling, roles as molecular motors, antibodies or enzymes[5].
Proteins are among the most complex molecules we know. Their sizes range from small peptides to huge macromolecules, consisting of several hundreds of amino acids stringed together. Each amino acid comprises two parts: the main chain, also called backbone, and the side chain. The main chain is the same for all amino acids, but they indeed differ in their side chain composition and so gain different physical and chemical properties. A variety of twenty different amino acids exist in the human body, of which the body itself can only produce twelve. The remaining eight amino acids are called essential and have to be supplied externally. The set of twenty amino acids can be roughly classified into three different groups: charged, polar and hydrophobic (Figure 2.1).

2.1.1. Primary Structure

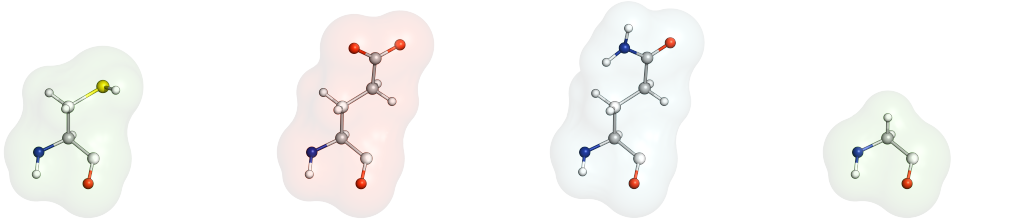
The sequence of amino acids in a protein is called the primary structure. It is a string composed of a twenty letter alphabet, where each letter represents a different amino acid. Two neighboring amino acids are bonded together through an amide bond, also referred to as the peptide bond. In this bond, both amino acids form a covalent bond between their main chain termini (Figure 2.2). For the formation, the OH-group of the N-terminal amino group creates a water molecule together with the H_N-atom of the C-terminal group, so that both main chains can form the peptide bond between their C and N atoms. The resulting partial double bond character of the peptide bond is caused by a delocalized free electron pair on the nitrogen atom N. It is also responsible for the planarity of the peptide plane, which can appear either in a cis- or in a trans-isomeric state. Aside from proline, all amino acids clearly favor occurring as a trans-isomer.

2.1.2. Secondary Structure

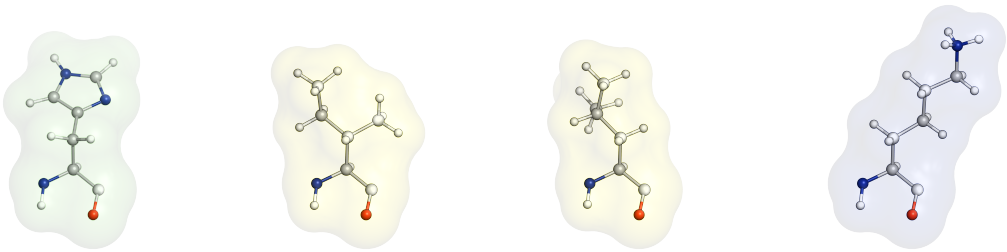
Due to the planarity of the peptide bond, the complexity of a $\mathbb{R}^3$ representation for a protein can be simplified to a reduced phase space described by all dihedral angles of the polypeptide (Figure 2.3). This causes the effective number of degrees of freedom to dramatically reduce to ≈ 5.5 per residue (for hemoglobins main chain[6]). The main chain angles Φ, Ψ and Ω are present in all amino acids and can be defined in the same way. Φ is defined by the atom sequence



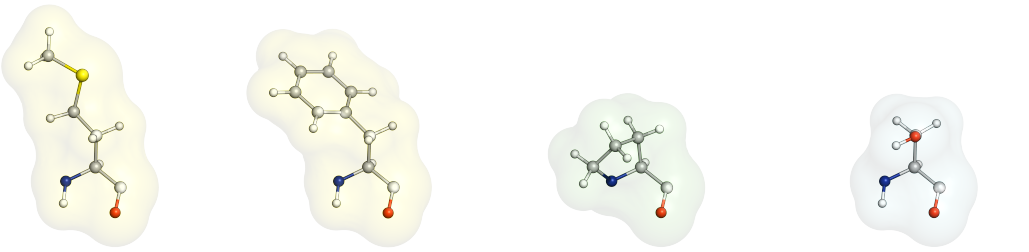
(a) *Alanine: ALA, A* **(b)** *Arginine: ARG, R* **(c)** *Asparagine: ASN, N* **(d)** *Aspartic acid: ASP, D*



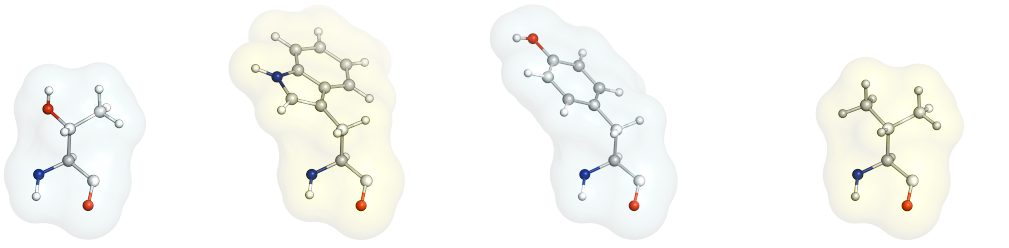
(e) *Cysteine: CYS, C* **(f)** *Glutamic acid: GLU, E* **(g)** *Glutamine: GLN, Q* **(h)** *Glycine: GLY, G*



(i) *Histidine: HIS, H* **(j)** *Isoleucine: ILE, I* **(k)** *Leucine: LEU, L* **(l)** *Lysine: LYS, K*



(m) *Methionine: MET, M* **(n)** *Phenylalanine: PHE, F* **(o)** *Proline: PRO, P* **(p)** *Serine: SER, S*



(q) *Threonine: THR, T* **(r)** *Tryptophan: TRP, W* **(s)** *Tyrosine: TYR, Y* **(t)** *Valine: VAL, V*

Figure 2.1.: *The 20 natural amino acids. Carbon atoms are shown in gray, Nitrogen atoms in blue, Oxygen atoms in red, Sulfur atoms in yellow and Hydrogen atoms in white. Additionally, the amino acids' surface is color encoded according to one of the five classes they belong to (**positive**: blue, **negative**: red, **polar**: cyan, **apolar**: yellow and **special**: green). This color code will be used for further representations of proteins.*

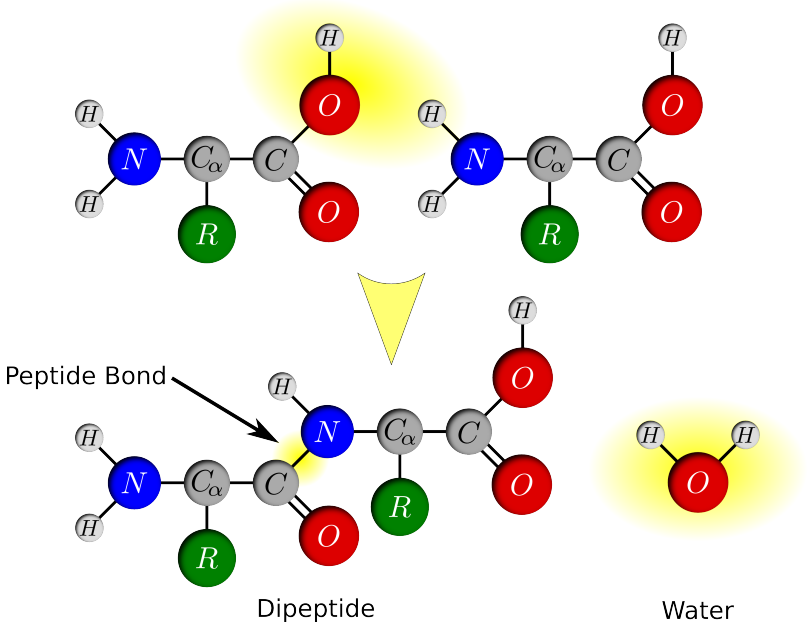


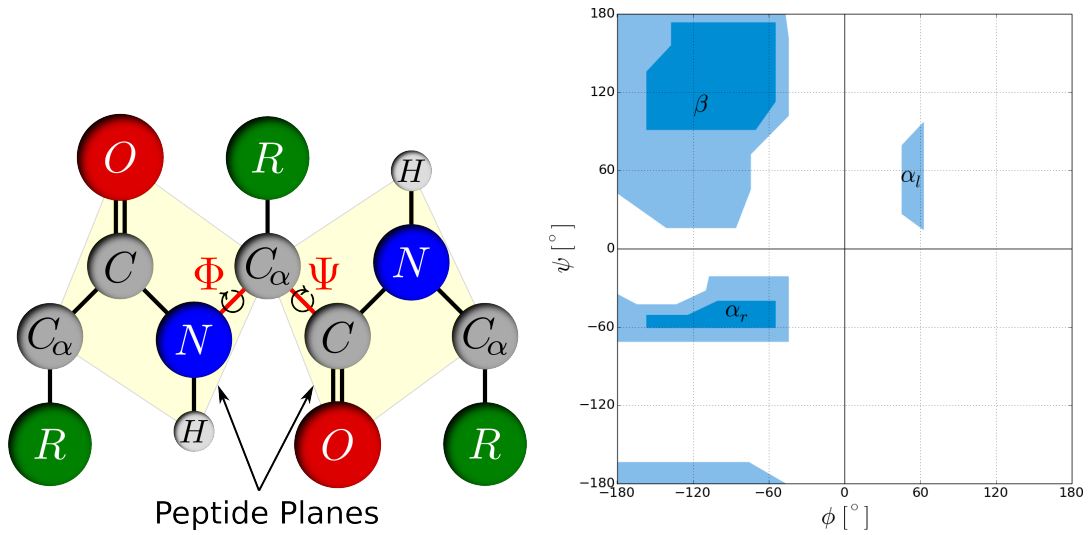
Figure 2.2.: *Two amino acids can form a covalent peptide bond between their terminal C and N atoms, which shows a partial double bond character responsible for the planarity of the peptide plane and can occur in either a cis or trans isomeric conformation. The condensation of multiple amino acids forms a polypeptide with a backbone build on a pattern of main chain atoms N - C_α - C.*

$C-N-C_\alpha-C$ and controls the $C-C$ distance, Ψ involves $N-C_\alpha-C-N$ and constrains the $N-N$ distance. The Ω dihedral restrains the two C_α atoms via $C_\alpha-C-N-C_\alpha$ and holds the peptide in either a trans-isomer ($\Omega \approx 180°$) or a cis-isomer ($\Omega \approx 0°$) state. The specification of the side chain dihedrals $\{\chi_i\}$ depends on the side chain composition.

The secondary structure is a local stable motif in the Φ-Ψ dihedral space. Pauling and Corey predicted a small set of possible motifs, but until now, only two of them could be discovered: the helix and the β-sheet conformation (Figure 2.4)[7–13]. Both motifs are stabilized by main chain hydrogen bonds. The most common helix, the α-helix, winds the main chain helically around itself and thereby forms hydrogen bonds between the CO-group of the i-th amino acid and the NH-group of the (i+4)-th amino acid. Theoretically, the sense of rotation could be left- as well as right-handed, but in nature, only the left-handed helix occurs, simply because of the side chain that then lies on the outside of the helix and so evades steric overlaps. The β-sheet consist of two or more polypeptide strands, stabilized via inter-strand hydrogen bonds. They can assemble either in a parallel or an anti-parallel manner However, the formation of a parallel β-sheet is more difficult because the single strands are more distant separated in sequence.

2.1.3. Tertiary Structure

The tertiary structure of a protein defines the three-dimensional arrangement of all secondary structure elements in its configuration. Many proteins spontaneously assume a unique tertiary structure. Based on Anfinsen's thermodynamic hypothesis, the native configuration is conformation with the lowest Gibbs free energy: $G = H - TS$, for the whole system[18]. In these cases the native structure of a protein is fully defined by its primary structure, the amino acid



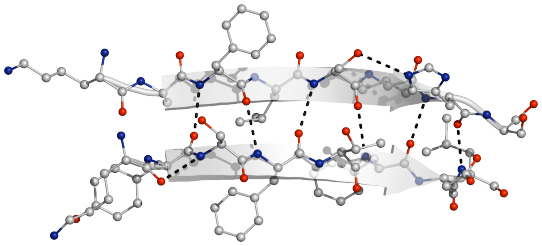
(a) *Mainchain dihedral angles.* **(b)** *Ramachandran plot.*

Figure 2.3.: *The dihedral degrees of freedom of a polypeptide comprise three main chain and various side chain dihedrals. Whereas* Ω *mostly remains constant at either* $0°$ *or* $180°$*, each* Φ *-* Ψ *pair can adopt numerous values (2.3a). However, steric exlcusion restricts these values to Ramachandran regions, shown in a ramachandran plot[14] (2.3b). Each region denotes the secondary structure composition (*α_r*,* α_l *or* β*) for the dihedral pair.*

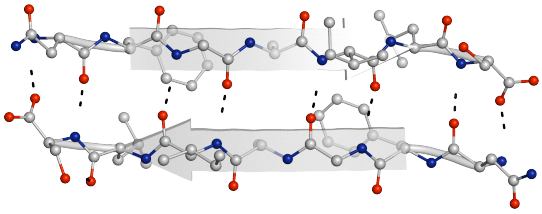
sequence, in a given environment. Anfinsen proved that unfolding is a reversible process for proteins. Once unfolded, many proteins are able to refold to their native structure. Finding this native structure within the giant conformational space of the protein in a random search process would take longer than the universe lifetime, while proteins fold on time scales of ms to s. To solve this folding paradox, also called Levinthal's paradox, folding pathways were postulated through which the folding process reaches the native state[19, 20]. The idea of single folding pathway was later extended to the existence of folding funnels. In this concept the free energy landscape is shaped like a multi dimensional funnel towards the global minimum, where the native conformation resides[21]. A schematic visualization of such a folding funnel is shown in Figure 2.5. Instead of describing a smooth funnel, for most proteins, the free energy landscape has a rough surface with numerous local minima. This makes protein structure prediction a challenging and computational expensive task[22, 23]. Nevertheless development of such methods is a worthwhile task, as the native configuration of a protein is experimentally relevant but often hard to obtain experimentally (Figure 2.6). Different theories exist how variations of the native state can lead to inactivation or malfunction of the protein, which may ultimately cause diseases. Many models have been proposed how the peptide chain finds its way down the funnel to the native structure:

- **Framework model**:
  The observation of some small proteins, where the tertiary structure denaturates before the secondary structure dissolves, suggested the introduction of the *framework model*[24]. It claims that in the folding process, the secondary structure is developed first, followed by the much slower process of tertiary structure formation[25, 26].

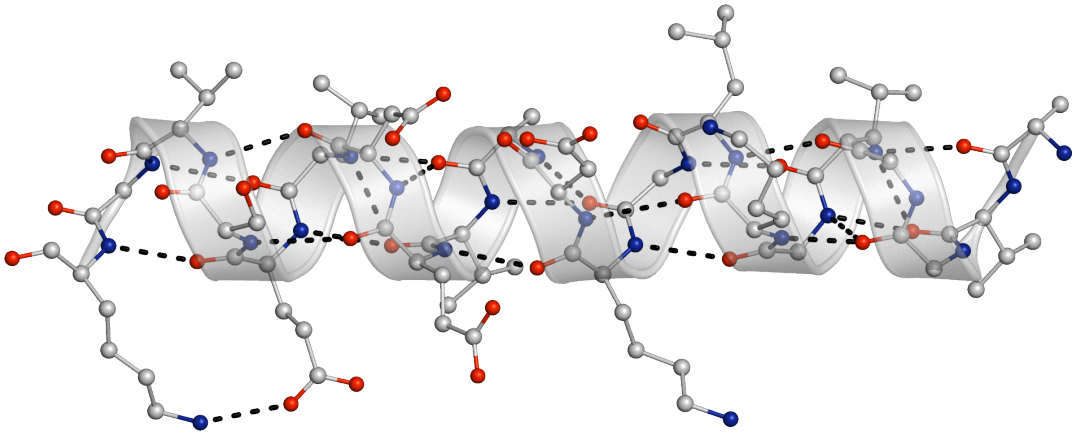
- **Hydrophobic collapse model**:



(a) *Parallel β-sheet, found in human amyloid polypeptides 2KIB[15]. The N-termini of both strands are aligned in the same direction to lie next to each other. Two residues with opposing C_α atoms do not share any hydrogen bond. Instead, the opposing bonds with the opposed residues successor or predecessor.*



(b) *Anti-parallel β-sheet, found in bovine pancreatic carboxypeptidase 1HDU[16]. Both strands are connected through a ladder of hydrogen bonds between each $C = O$-group on the one and each $N - H$-group on the opposing side.*



(c) *Part of a Dermcidin α-helix, found in a hexameric anti-microbial peptide channel 2YMK[17]. Each $N - H$-group of residue n forms a hydrogen bond with the $C = O$-group of the $(n + 4)$-th amino acid. By orienting all $N - H$-groups to face the N-terminus and all $C = O$-groups facing the C-terminus, the helix gains a polarity.*

Figure 2.4.: *Secondary structure elements are local stable motifs, dominantly shaping the appearance of proteinous structures (2.4a, 2.4b and 2.4c). Stabilized through characteristic hydrogen bond patterns, denoted by black dashed lines, 2^{nd} structure elements effect protein's potential energy.*

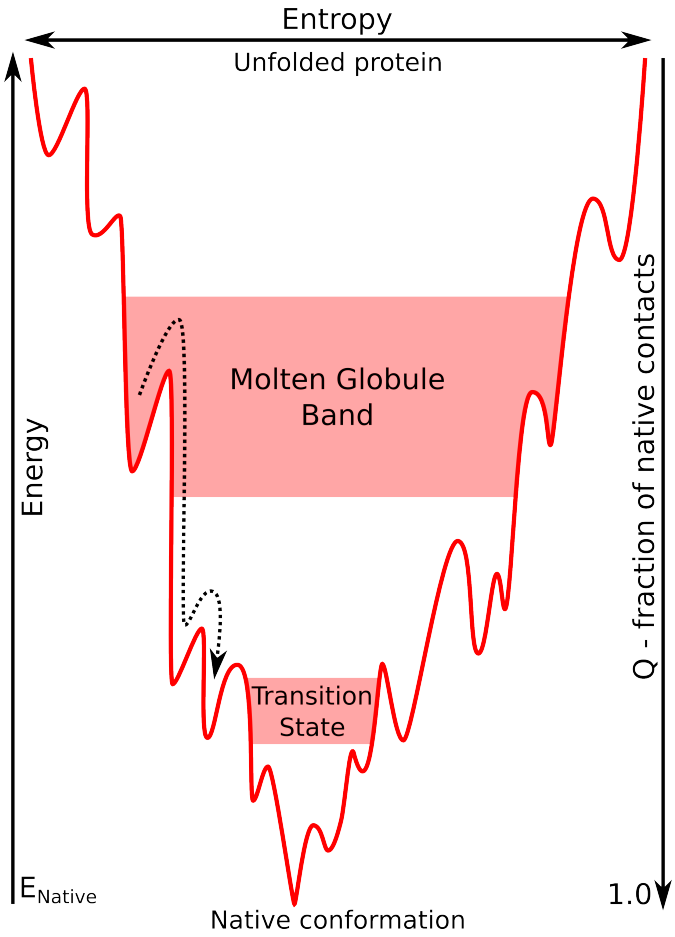
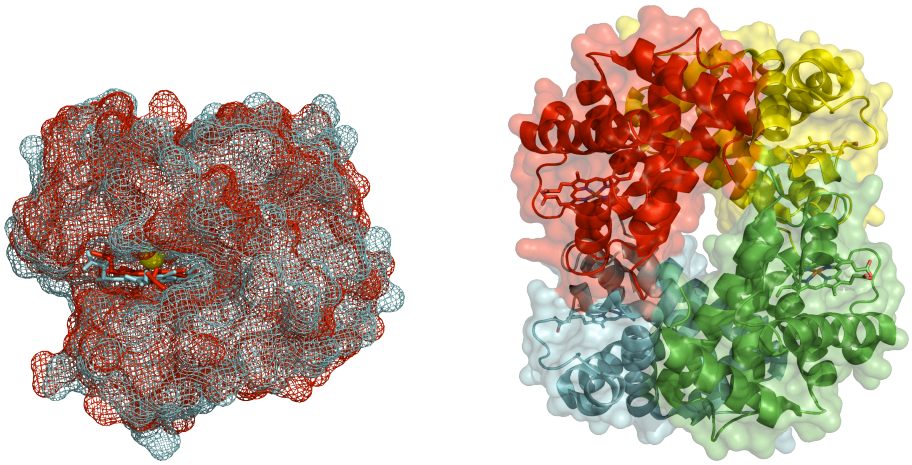


Figure 2.5.: *The introduction of folding funnels extended the idea of folding pathways to a more complex view. Whereas the native conformation displays the minimum of the free energy E_{native} and lies at the maximal fraction of native contacts Q, the unfolded ensemble is completely dominated by entropy. The transition from unfolded to folded state may has to overcome several local minima, pass the molten globule band to finally reach the transition state from which the energy decreases monotonously reaching either the native state or a possible substate. The figure is based on Onuchics work[21].*

Globular proteins have a large hydrophobic core[27]. For many proteins this hydrophobic core develops in the early folding stage. The *hydrophobic collapse model* proposes the rapid collapse of the polypeptide chain into such an unstructured non-native globule without the presence of secondary structure elements[28]. Afterwards the slow search for the native conformation is initiated[29].

- **Nucleation condensation model**:
  Several globular proteins possess distinct structural regions in their peptide chain. In the *nucleation condensation model*, these regions can independently form some structure in the early stage of folding, which are then condensed into the full tertiary structure[30].

In addition to proteins showing unambiguous folding behavior, often the mixture of all three models represents the most accurate model.



(a) *HEM group binding.* **(b)** *Hemoglobin complex.*

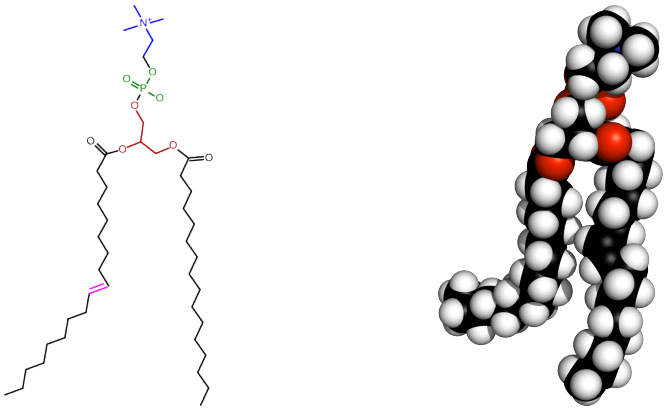
Figure 2.6.: *The hemoglobin protein consists of four independent subunits, each one carries an iron-containing heme-group able to bind oxygen. It is responsible for the oxygen transport in red blood cells of all vertebrates. The binding of a heme-group invokes large conformational changes: red↔blue (2.6a)[31]. An assembled hemoglobin protein is able to transport four O_2 molecules at once (2.6b)[32, 33].*

2.2. Biological Membranes

Membranes are essential for the life of all cells. Every living cell is surrounded by a biologic membrane that delimitates the cytosol from the extracellular environment. In spite of their diverse functionality, e.g. ion gradients, ATP synthesis or signal transcription, they all are composed based on the same framework. Lipids, small, hydrophatic organic molecules (Figure 2.7), form this basis and make up to 50% of the mass for most animal cell membranes. The remainder is almost solely composed of membrane proteins.
The most common lipid components found in membranes are phospholipids. These lipids consist of a polar head group and two hydrophobic hydrocarbon tails. Although their head groups may differ over various phospholipids and may be charged (e.g. phosphatidylglycerol and phosphatidylserine) or uncharged (e.g. phosphatidylcholine and phosphatidylethanolamine), they all assemble spontaneously into lipid bilayers with a thickness of $\approx 5nm$ (Figure 2.8).
Those bilayers form an impenetrable barrier for the majority of water soluble molecules. Here, the tails of both lipid layers form a hydrophobic core, shielded from the surrounding water by the hydrophilic head groups. The cytosolic face of the membrane points to the inner of the cell, the cytoplasm. The exoplasmic face defines the shell of the cell and points away from the cytosol[5, 34].

2.2.1. Membrane Proteins

While lipids are responsible for the formation of the lipid bilayer and therefore describe the basic structure of a biomembrane, proteins take care for the bulk of membrane functions. To that effect, there is a huge variety of membrane proteins. Some of them span the whole lipid bilayer, others are bound to one side of the membrane; some can diffuse freely through the



(a) *POPC 2D.* **(b)** *POPC 3D.*

Figure 2.7.: *Structural constitution of phospholipids using the example of phosphatidylcholines lipids (POPC). The hydrophilic head group attaches two long hydrophobic hydrocarbon tails at the clycerol (green) which is bound to choline (blue) via a phosphate (red). Usually, one of the fatty acid comprises one or more cis-double bonds (pink), which are responsible for the small kink in the tail.*

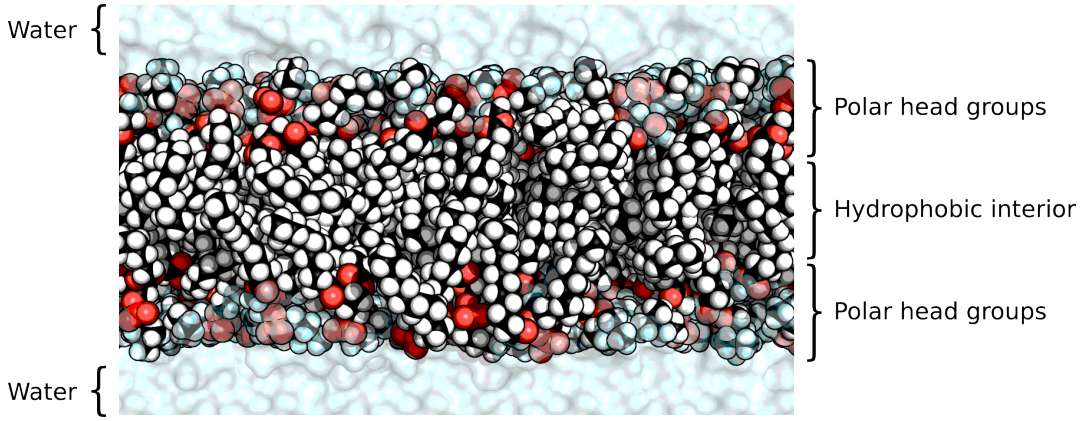


Figure 2.8.: *The highly flexible lipids (DLPC) spontaneously form lipid bilayers. Whereas the lipids polar head groups, describing the membrane surfaces, face the surrounding water molecules, a cross section through the membrane shows that their hydrophobic fatty acids lie shielded within the membrane and form its hydrophobic interior.*

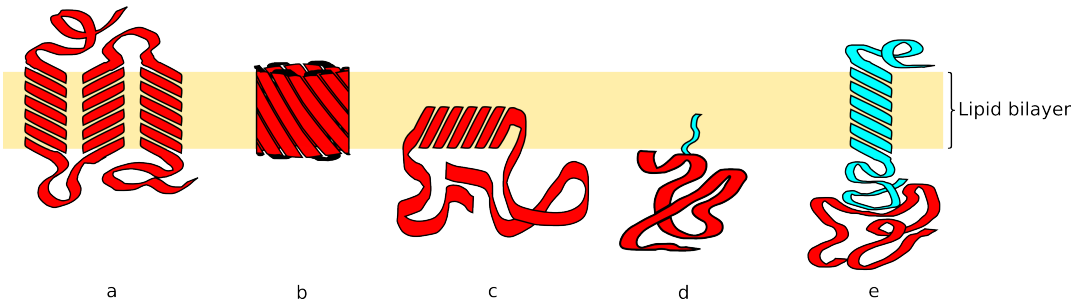


Figure 2.9.: *Proteins can associate with lipid bilayers in various ways. Most membrane proteins form helical structures (a), others assemble into complete beta barrels spanning the membrane (b), some proteins only attach to either one side of the bilayer having amphipatic structural parts (c), others bind indirectly to the membrane via covalent bound molecules (d) or they can be attached non-covalently to other membrane proteins (e). Figure based on Alberts work[5].*

membrane, others are constrained to a specific domain. The proteins association with the lipid bilayer is strongly reflected in its functionality (Figure 2.9). Basically one distinguishes between two different membrane protein types:

- **Transmembrane proteins**:
  A *transmembrane protein* spans the whole bilayer and is accessible at both membrane sides at once. Like lipids, they are amphipathic molecules having both, hydrophobic and hydrophilic regions. While the hydrophobic parts of the protein rest in the inner region of the membrane and bind to the lipid tails, the hydrophilic parts lie, exposed to the water, on both membrane surfaces.

- **Peripheral membrane proteins**:
  Some membrane proteins do not lie in contact with the hydrophobic part of the bilayer and are bound to the membrane via further molecules or membrane proteins.

Many functional complexes in membranes consist of various subunits of membrane proteins. The first experimental resolved complex is the bacterial photosynthetic reaction center [35], composed of four distinct subunits (Figure 2.10).

2.2.2. Membrane Transport

Lipid bilayers are highly impermeable for polar molecules. Their non-zero charge and the surrounding water shell inhibit the penetration into the hydrocarbon phase of the bilayer. Nevertheless, to perform vital transport processes like nutrient uptake, metabolism or regulation of ion concentration, cells are using specialized transport proteins. Between 15% and 30% of a cells genes are used to code for transport proteins. The transport process is highly specific and each transporter is only functional for a certain molecule or class of molecules.

Membrane transporters can be grouped into two main classes (Figure 2.11):

- **Carriers** bind to a specific solute and have to undergo several conformational changes in order to transport the solute across the membrane.

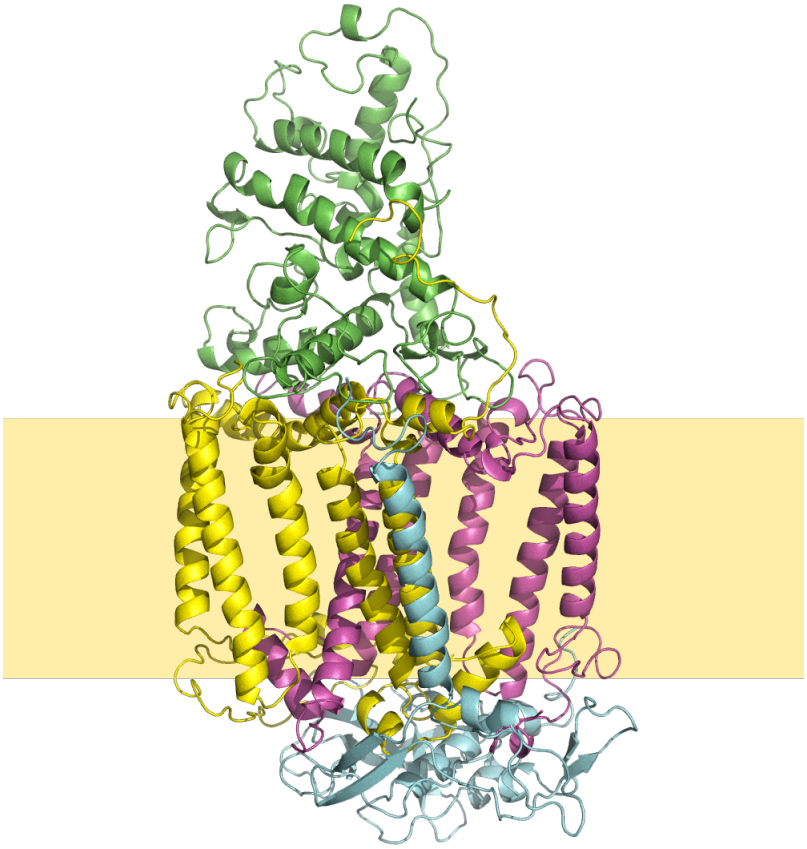


Figure 2.10.: *The bacterial photosynthetic reaction center was the first experimental resolved complex[35]. It comprises four distinct subunits and spans the membrane to catalyze the first steps during photosynthesis. After the absorption of photon-energy the electrons were guided fast across the photosynthetic membrane.*

(a) *Carriers.* **(b)** *Channels.*

Figure 2.11.: *The membrane transporter classes comprise carriers (2.11a) which undergo serveral conformational changes in order to transport a solute across the membrane and channels (2.11b) which only interact weakly with the solutes and allow a faster traversal[5].*

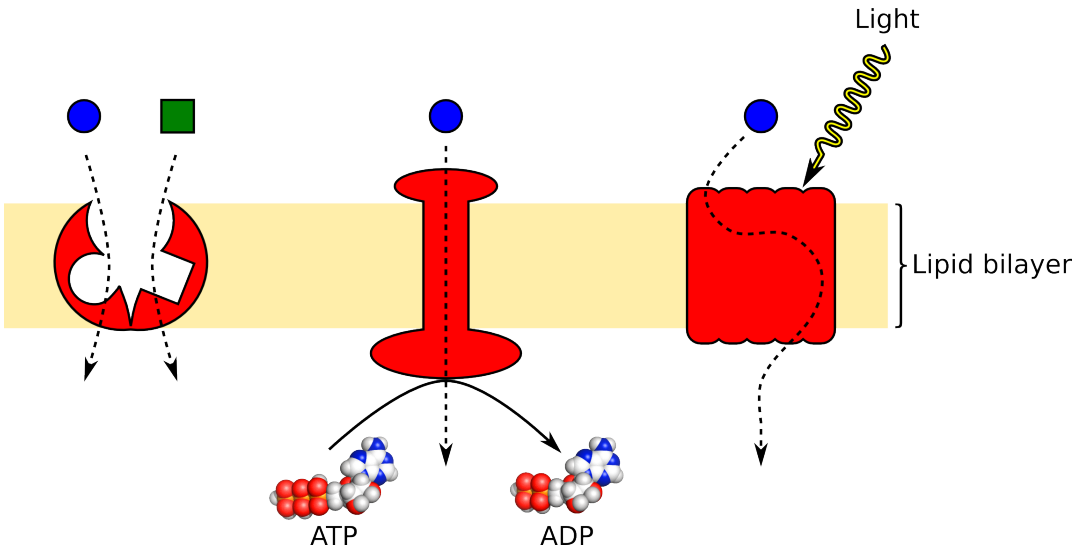


Figure 2.12.: *The active membrane transport requires an external supplied energy source. It can be either coupled with a second, inverse directed transport gradient (left), energetically driven by the hydrolysis of ATP (middle) or light-driven by an external light source (right). Figure based on Alberts work[5].*

- **Channels** develop aqueous pores which allow specific solutes to traverse by only weakly interacting with them. This transport mechanism is way faster than the transport via carriers.

A transport process is called **passive**, if it is driven by either a concentration gradient or by an electrochemical gradient for uncharged or charged solute, respectively. An **active** transport is directed against such a gradient with the need of having an additional energy source applied. The three main mechanisms for supplying the additional potential to facilitate an active carrier transport are (Figure 2.12):

- **Coupled carriers** join the transport with a second, opposed directed, transport along the gradient.
- **ATP hydrolysis pumps** couple the flow against the gradient with the hydrolysis of ATP.
- **Light driven pumps** are mostly found in bacteria and archaea. The required energy is supplied by an external light source. The most prominent example is the bacteriorhodopsin[36, 37].

Most channel proteins in the plasma membrane build highly selective pores, able to open and close on short time scales. Due to their involvement in the transport of organic ions, they

are referred to as **ion channels**. In comparison to carrier proteins, ion channels can guide ions 10^5 times faster. Two important properties distinguish ion channels from simple aqueous pores:

Ion channels ...

- ... offer an **ion selectivity** that allows them to only let pass certain organic ions. These ions are separated from bonded water molecules and screened by the channels selectivity filter what emerges as the rate limiting step in the transport process.
- ... are **gated**, what allows them to open and close their transition pathway based on various stimuli:
  - *voltage-gated*: voltage changes of the membrane potential
  - *mechanically gated*: a mechanical stimulus
  - *ligand-gated*: binding of a ligand molecule:
    - *transmitter-gated*: an extracellular mediator (neurotransmitter)
    - *ion-gated*: an intracellular mediator (ion)
    - *nucleotide-gated*: a nucleotide

2.3. Thermodynamic Ensemble Techniques

With always growing advancements in computational science, the theoretical description of molecular systems becomes more and more significant. Where in the beginning of the computer based description of proteinous systems on an atomistic level the investigated systems consisted of only a handful of atoms[38–42] to describe the behavior of basic liquids, current simulations can describe billions of particles at once[43, 44] or large biological complexes comprising several millions of atoms[45–48]. However, this progress strongly correlates with the development in the computer field, whereas the fundamental methods for the description of molecular systems remained unchanged. Molecular simulations are capable to realistically describe thermodynamic processes either based on the solution of Newtons equations of motion or on stochastically generated Markov chains.

2.3.1. Molecular Dynamics

For the investigation of many biological processes, a description in a classical force field is sufficiently accurate and can be described by solving Newtons equations of motion (2.1). Molecular dynamics (MD) methods represent a deterministic way for iteratively solving these equations

numerically by integrating over a discretized time step Δt[39].

$$m_i \frac{\partial^2 \vec{x_i}}{\partial t^2} = \vec{F_i}, \qquad i = 1, ..., N \tag{2.1}$$
$$\vec{F_i} = -\frac{\partial V}{\partial \vec{x}_i}$$

The step-wise integration and data collection leads to a series of snapshots holding the positions and velocities of the simulated particles during a period of time. Observables can be measured as a time average of a function of the recorded configurations.
One of the most common integration techniques is the application of a Taylor polynomial for the initial step:

$$\vec{x}(t+\Delta t) = \vec{x}(t) + \frac{\vec{p}(t)}{\vec{m}}\Delta t + \frac{\dot{\vec{p}}(t)}{2\vec{m}}\Delta t^2 + \mathcal{O}(\Delta t^3)$$
$$\vec{p}(t+\Delta t) = \vec{p}(t) + \dot{\vec{p}}(t)\Delta t + \frac{\ddot{\vec{p}}(t)}{2}\Delta t^2 + \mathcal{O}(\Delta t^3) \qquad ,$$

followed by a Verlet-Störmer integrator for all further steps[49]. It is based on the observation that:

$$\vec{x}(t+dt) + \vec{x}(t-dt) = 2\vec{x}(t) + \frac{\dot{\vec{p}}(t)}{\vec{m}}dt^2 + \mathcal{O}(dt^4)$$
$$\vec{x}(t+dt) - \vec{x}(t-dt) = 2\frac{\vec{p}(t)}{\vec{m}}dt + \mathcal{O}(dt^3) \qquad ,$$

what can be rewritten for a small time increment Δt to:

$$\vec{x}(t+\Delta t) = 2\vec{x}(t) - \vec{x}(t-\Delta t) - \frac{1}{\vec{m}}\left.\frac{\partial H}{\partial \vec{x}}\right|_t (\Delta t)^2$$
$$\vec{p}(t+\Delta t) = \vec{m}\frac{\vec{x}(t+\Delta t) - \vec{x}(t-\Delta t)}{2\Delta t} \qquad .$$

An equivalent algorithm, in terms of giving the identical trajectories, is the leap-frog[50] method which updates the momentum at half-time $t + \frac{\Delta t}{2}$:

$$\vec{x}(t+\Delta t) = \vec{x}(t) + \Delta t\frac{\vec{p}\left(t+\frac{\Delta t}{2}\right)}{\vec{m}}$$
$$\vec{p}\left(t+\frac{\Delta t}{2}\right) = \vec{p}\left(t-\frac{\Delta t}{2}\right) + \left.\frac{\partial H}{\partial \vec{x}}\right|_t \Delta t \qquad .$$

Both, the Verlet-Störmer and the leap-frog methods are volume preserving methods:

$$|V(t)| \equiv \int\int_{V(t)} d\vec{x}\, d\vec{p} = \int\int_{V(0)} d\vec{x}\, d\vec{p} = |V(0)| \qquad .$$

The systems kinetic energy k and total energy H are given by:

$$k(\vec{p}) = \frac{1}{2}\left\|\frac{\vec{p}}{\sqrt{\vec{m}}}\right\|^2, \qquad \text{with} \quad \frac{\vec{p}}{\sqrt{\vec{m}}} = \left(\frac{p_1}{\sqrt{m_1}}, ..., \frac{p_N}{\sqrt{m_N}}\right)$$
$$H(\vec{x}, \vec{p}) = U(\vec{x}) + k(\vec{p}) \qquad .$$

For a closed system, the law of conservation of energy grants the constancy of the total energy. Depending on the simulated ensemble, the system may also be coupled with thermostats or barostats to ensure that the environmental conditions remain at the required values[51].

An enormous disadvantage of the MD method is the magnitude of the integration step. To guarantee the robustness of a simulation, the integration step Δt has to be in the order of the fastest occurring transition. Therefore it is restricted to the time scale of atomic vibrations which take place in the $1fs$ time scale. This is way below the time scale range of μs up to s where most of the relevant processes take place. The investigation of such processes using molecular dynamics may be either not feasible or it requires a massive computational effort.

2.3.2. Markov Chains

Markov chains[52] are memoryless, random processes traversing a finite or countable number of states. A random process $(X_0, X_1, ...)$ in the state space $S = \{s_1, ..., s_k\}$ is called a Markov chain with transition matrix P if:

$$\begin{aligned} &P\left(X_{n+1} = s_j | X_0 = s_{i_0}, X_1 = s_{i_1}, ..., X_{n-1} = s_{i_{n-1}}, X_n = s_i\right) \\ &\quad = P\left(X_{n+1} = s_j | X_n = s_i\right) \\ &\quad = P_{i,j} \qquad , \end{aligned}$$

with $i, j \in \{1, ..., k\}$ and all $i_0, ..., i_{n-1} \in \{1, ..., k\}$.

For an irreducible and aperiodic Markov chain, π is the stationary distribution if it is reversible:

$$\pi_i P_{i,j} = \pi_j P_{j,i} \qquad .$$

Starting from an arbitrary initial distribution, the Markov chain convergence theorem guarantees that the chains distribution converges to π at time n, for $n \to \infty$[53]. For a non-countable state space, one has to introduce the Markov core as generalization to the transition matrix P.

Metropolis Hastings

Monte Carlo methods are widely used to perform simulations from a complex probability distribution $\pi(\vec{x})$. The Metropolis Hastings (MH) algorithm is a well known method for Bayesian estimation which allows the generation of reversible Markov chains having a stationary distribution that is equal to the target distribution $\pi(\vec{x})$. It was proposed 1953 by Metropolis et. al[54] and modified 1970 by Hastings[55].

Assuming the system currently resides in the state $\vec{x}_t$. Then the MH allows one to use an arbitrary Markov transition function $T(\vec{x}_t, \cdot)$ to generate a new proposal $\vec{y}$ and accept it according to the

probability:

$$\alpha_{MH} = \min\left\{1, \frac{\pi(\vec{y})T(\vec{y},\vec{x}_t)}{\pi(\vec{x}_t)T(\vec{x}_t,\vec{y})}\right\} \quad .$$

The stepwise appliance breaks down complex tasks into an ergodic Markov chain containing feasible small pieces, which is reversible because it satisfies the *detailed balance* condition:

$$\pi(\vec{y})P(\vec{y},\vec{x}) = \pi(\vec{x})P(\vec{x},\vec{y}) \quad ,$$

where $P(\vec{x},\vec{y})$ denotes the transition kernel from $\vec{y}$ to $\vec{x}$[56].

The ergodicity theorem says that in a large system, the MD *time averages* will converge against the MH *configuration averages*[57]:

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t h(\vec{x}_s)\,ds = \frac{1}{Z}\int h(\vec{x})e^{-\frac{U(\vec{x})}{\beta T}}\,d\vec{x} \quad .$$

2.4. Membrane Simulations

Approximately a quarter of an eukaryotic cells genome codes for membrane proteins and about 50% of the available drug targets are membrane proteins. Nevertheless, only a few hundred distinct high-quality structures for membrane proteins are currently known. The complex lipid bilayer which forms a two- dimensional liquid crystalline system hinders the experimentalists from explicitly resolving protein-membrane interactions but makes them a special point of interest for computer simulations[58]. Using micelle forming detergents, to which peptides can bind, liquid state NMR can be applied to resolve all structural motifs[59]. But the environmental change influences the peptides structure and may lead to measurements of non-native structures. Nonetheless, to investigate the proteins behavior within its native membrane environment, simulation approaches could play a key role[60].
An example is the simulation of mechanosensitive channels discovered spontaneous structural changes; starting from the open-state crystal structure without additional restraints to the backbone, the transmembrane channel closed itself spontaneously and could be widened again, by applying a surface tension[61, 62]. However, these simulations are often restricted to small time scales due to the numerous count of involved degrees of freedom; but even the most basic interactions occur on a microsecond time scale. Thus, additional to an atomistic description, two further approaches for characterizing protein-membrane interactions have been evolved:

- **Explicit** membrane simulations treat the whole environment explicitly, meaning molecular water (TIP3P/TIP4P[63], SPC[64], SPC/E[65]) and atomistic lipid bilayers[66] are included.

- **Implicit** membrane descriptions approximate the interactions between the protein and the surrounding environment (e.g. water or lipids) in an implicit manner, using a continuum electrostatic ansatz.

- **Coarse-grained** membranes reduce the number of degrees of freedom by introducing beads, replacing complete groups of atoms. These beads are restrained by artificial interactions to reflect realistic thermodynamic properties of the complex.

2.4.1. Explicit Membranes

The immense quantity of involved atoms in combination with the femtosecond time step of an MD simulation, currently restricts these simulations to realms far below the microsecond time scale, where the first simple protein-membrane interactions occur. Nonetheless, an explicit description based on transferable parameters makes a reliable model for the quantitative prediction of kinetic and thermodynamic properties. In this process, explicit lipid molecules seem to have wide influence on the described protein and its functionality[67].
Using an united-atom parametrization for the bilayer forming lipids, microsecond MD simulations could show the formation of native transmembrane helices in WALP peptides[68]. Various in situ MD simulations of ion channels, selective/non-selective channels and membrane proteins produced great results and granted insights into previous hidden systems[60, 69]. Also, MD simulations of explicit lipid bilayers enriched with diverse ions concentrations, reproduced accurate data, staying in well agreement with the Gouy-Chapman theory of charged membranes[70].

2.4.2. Implicit Membranes

Implicit solvent methods try to average over the solvents degrees of freedom to approximate the solute as potential of mean force[71]. The electrostatic free energy of a protein related to a membrane can be expressed as a sum over two terms, a non-polar part and electrostatic contributions[72]. The polar part of such a continuum model can be calculated by solving the Poisson equation[73]:

$$\nabla \left[\epsilon(\vec{r})\nabla\phi(\vec{r})\right] = -4\pi\rho(\vec{r}) \qquad ,$$

where $\phi(\vec{r})$ is the electrostatic potential, $\rho(\vec{r})$ is the charge distribution and $\epsilon(\vec{r})$ is the dielectric constant. However, due to the high costs of numerical solving this problem[74] by calculating the solvent-induced reaction field energy:

$$\Delta G^{elec}_{\epsilon_p \to \epsilon_w} = G^{elec}_{\epsilon_w} - G^{elec}_{\epsilon_p} \qquad ,$$

where ϵ_p and ϵ_w are the dielectric constants for the protein and solvent, respectively, via solving the integral for the electrostatic energy:

$$G^{elec} = \frac{1}{2}\int_V \rho(\vec{r})\phi(\vec{r})\, dV \qquad ,$$

this is not suitable for a dynamic simulation.

The most prominent method for approximately solving this equation with high accuracy, is the

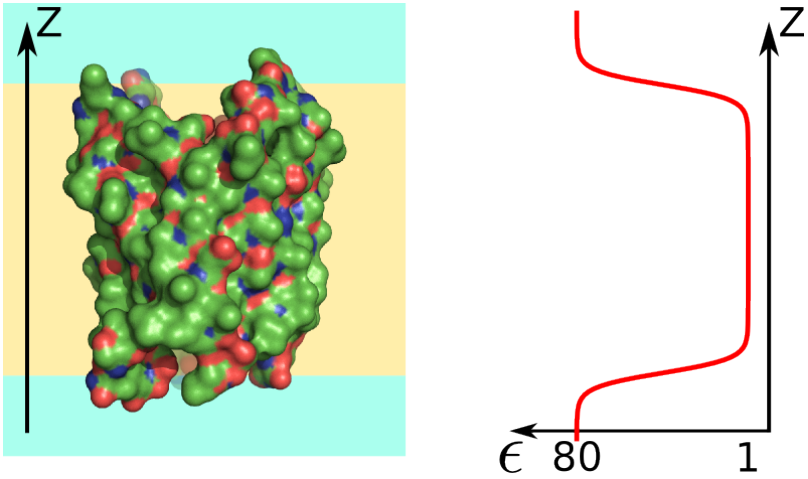


Figure 2.13.: *Implicit description of a heterogeneous lipid bilayer (left). The dielectric constant ϵ is shown along the normal of the implicit membrane plane (right). Such a distribution is used by Generalized Born approaches to approximate the polar part of the solvation free energy for implicit membrane calculations.*

Generalized Born (GB) formalism[75]:

$$\Delta G^{elec}_{\epsilon_p \to \epsilon_w} = -\frac{1}{2}\left(\frac{1}{\epsilon_p} - \frac{1}{\epsilon_w}\right)\sum_{i,j}\frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i R_j \exp\left(-r_{ij}^2/4R_i R_j\right)}} \quad ,$$

where R_i are the Born radii which can be calculated using different approaches[76, 77] and have a crucial influence on the resulting accuracy[78]. Modern GB solvers can approximate the polar part of the solvation free energy within an error of $\approx 1\%$[79, 80].

The simulation, using implicit electrostatics within a GB model, is limited to homogeneous environments like soluted proteins, but the implicit handling of membranes requires a heterogeneous model, describing the interior of a membrane with a low dielectric constant and the lipid headgroup regions as well as the water with a high ϵ (Figure 2.13)[81]. Therefore the GB formalism has been extended, to allow the implicit treatment of membranes[82–84] and recently also supports the deformation of the lipid layer[85]. It has been shown that the implicit formulation of protein-membranes complexes within a GB model produces reasonable results which compare well with explicit calculations and also with experimental data[86].

2.4.3. Coarse-Grained Membranes

Coarse-graining (CG) effectively reduces the number of degrees of freedom of the simulated system by discarding the atomistic handling of the structure and replacing it with the introduction of CG beads. Each single bead represents $2-5$ heavy atoms and interacts with all other CG beads via artificial bonded and non-bonded interactions (Figure 2.14) [58]. The parametrization is chosen to reflect realistic thermodynamic properties. The removal of degrees of freedom also eliminates the fastest transition in the system, the atomic vibration. This allows the use of much larger integration time steps around $40fs$. Together with the benefit from the less computational effort, a CG simulation can provide speedups of $10^2 - 10^3$.

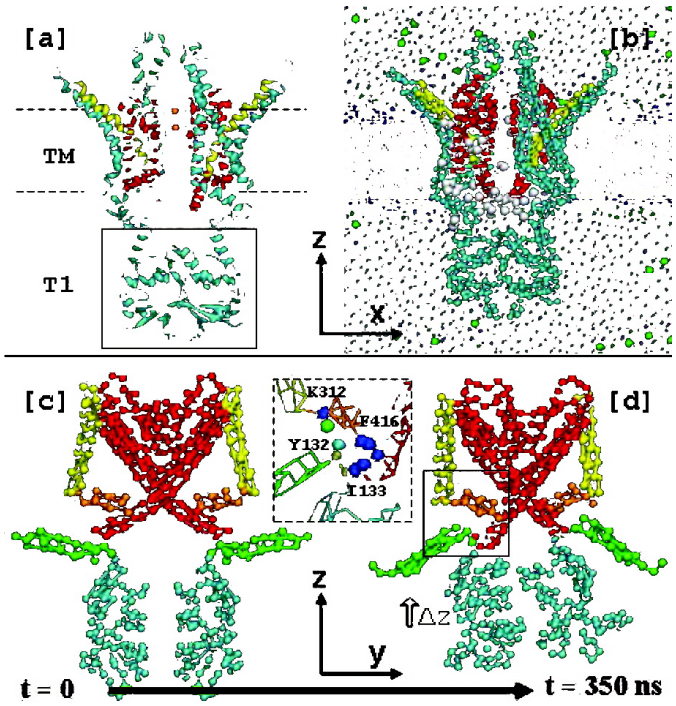


Figure 2.14.: *Coarse-grained representation of a protein-membrane complex. Coarse graining makes the simulation of large protein complexes feasible. Each bead represents* $2 - 5$ *heavy atoms; artificial interactions allow the beads to interact and reflect physical properties of the underlying atomic system. Reprinted with permission from[92]. Copyright © (2008) American Chemical Society.*

This enhanced simulation speed allows existing CG models[87–90] to produce trajectories reaching the microsecond time scale[91].

3. SIMONA - Simulation of Molecular and Nanoscale Systems

3.1. Motivation

Increasing computational resources increased the importance of molecular simulation methods for the understanding of (bio-)molecular functionality in the last decades[93–95]. In addition to the enhancements of theoretical methods, also experimental techniques have been developed or improved to resolve smaller and smaller scales [96, 97]. Monte Carlo simulation techniques, which have been a useful toolkit for describing thermodynamic properties in the field of condensed matter physics[98], more and more lose their importance compared to the flood of available molecular dynamics methods and techniques[99–106]. Well supported simulation packages like GROMACS[107], CHARMM[108], AMBER[109], NAMD[110], or LAMMPS[111], to name only the most popular ones, have been established with the ability to perform simulations of moderate sized systems for several hundreds of microseconds [95, 112, 113] or even giant systems on shorter time scales[114, 115]. The fact that MD simulations are limited by integration time step in the femtosecond range for biological systems still prevents this method from being able to describe long time scale processes with adequate statistical data. This has caused the development of many simulation strategies like replica exchange methods (RE) [116, 117], umbrella sampling (US)[118] or coarse graining (CG)[88, 90]. Most of them enhance the sampling of the systems conformational space by increasing the transition rates (RE), lowering barriers (US) or simply reducing its dimensionality (CG) with the consequences of losing either the kinetic or the structural information.
However, there are plenty of processes where the thermodynamic information alone is sufficient to describe or guide an experiment or even predict its outcome. These processes are distributed over many disciplines, ranging from biomolecular systems (e.g. protein stability, protein-protein interaction, or protein-ligand interaction) to material science (e.g. aggregation, growth of nanoparticles or determination of properties of thin films). For all these processes a MC based method could be able to outperform MD approaches.
I have participated in the development of SIMONA (**SI**mulation of **MO**lecular and **NA**noscale systems) to be a generic MC-based simulation package, which enables rapid prototyping of both: force fields and simulation methods for efficient MC simulations[2] (Figure 3.1).

3.2. Force Fields

A force field is a collection of several force field terms, each describing an abstraction of a distinct property of the simulated system. Accumulated, they result in an approximation for the potential energy of the system. Many force fields have been developed for proteinous systems (e.g. GROMOS[119, 120], CHARMM[121–123], AMBER[109, 124–128], PFF[129, 130]) or

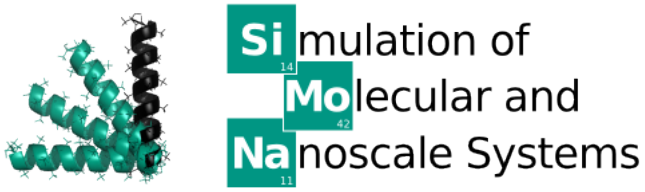


Figure 3.1.: *SIMONA[2] is a generic MC simulation toolkit, free for academic use and can be downloaded at* `http://www.int.kit.edu/nanosim/simona`.

generic molecular systems (e.g. OPLS [131]) providing atomic parameters like partial charges, atom masses or covalent bond definitions as well as force field terms.
SIMONA supports the one-click parametrization of a proteinous structure in either one of the PFF0x[129, 130] force fields, in one of the AMBER99SB[128] force fields treating the environment implicitly by using an acccurat and efficient Generalized Born approach[77] or by manual combining single force field terms. The force field terms currently implemented in SIMONA can be roughly divided into three potential classes:

- **Non-bonded interactions** are calculated based on a neighbor list containing non-bonded atoms within a certain radius.
- **Bonded interactions** are used to describe energies involving covalent bonded atoms like bond-stretching, angle-bending or dihedral-rotation.
- **Restraints** are used to add arbitrary potentials over arbitrary properties. (e.g. position restraint, distance restraints or secondary structure restraints)

3.2.1. Non-bonded Interactions

Non-bonded interactions are pair-additive force field terms between two atoms, which do not share a covalent bond, of the form:

$$U_{FT}(\vec{x}) = \sum_{i=0}^{N} \sum_{j=0, j\neq i}^{N} F(\vec{x}_i, \vec{x}_j) \quad . \tag{3.1}$$

Analog to the Einstein notation, in the following the $\sum$ will be omitted and implied by a twice appearing index variable, excluding the diagonal for $\mathcal{O}(N^2)$ terms[132].

Lennard-Jones Potential

The Lennard-Jones potential comprises a description for two essential physical properties, the Pauli exclusion, modeled by an repulsive $\sim r^{-12}$ term, and the dipole-induced London attraction[133]. Repulsive electrons between interacting molecules invoke the formation of instantaneous dipoles. The resulting dispersion forces can be described by an attractive potential that

decays with $\sim -r^{-6}$. The most common expression for the Lennard-Jones potential is:

$$U_{LJ}(\vec{x}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad . \tag{3.2}$$

The depth of the Lennard-Jones potential is controlled via $\epsilon_{ij} = \sqrt{\epsilon_i \cdot \epsilon_j}$ and the equilibrium distance $r_m = 2^{1/6}\sigma_{ij}$ defines its global minimum, with $\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2}$. The two parameter ϵ and σ are called Lennard-Jones parameters and are unique for each atom type.

Coulomb Electrostatics

Electrostatic interactions between two charged particles with partial charges q_i and q_j are often described using partial charge Coulomb electrostatics:

$$U_{Coulomb}(\vec{x}) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r}\frac{1}{r_{ij}} \quad . \tag{3.3}$$

With the vacuum permittivity $\epsilon_0 = 8.8541878\frac{C^2}{Nm^2}$ and the relative dielectric constant $\epsilon_r = 1$.

Implicit Solvent

Analog to the GB formalism used to describe implicit membranes in section 2.4.2, SIMONA supports the evaluation of the aqueous environment implicitly by describing it as a dielectric continuum. The environmental influences on the protein is divided into a surface proportional, non-polar part ΔG_{np} and a polar part $\Delta G^{elec}_{\epsilon_p \to \epsilon_w}$:

$$\Delta G_{\text{Solvent}} = \Delta G_{np} + \Delta G^{elec}_{\epsilon_p \to \epsilon_w} \quad ,$$

with:

$$\Delta G_{np} = \gamma \sum_i S_i \quad \text{, and}$$

$$\Delta G^{elec}_{\epsilon_p \to \epsilon_w} = -\frac{1}{2}\left(\frac{1}{\epsilon_p} - \frac{1}{\epsilon_w}\right)\sum_{i,j}\frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i R_j \exp\left(-r_{ij}^2/4R_iR_j\right)}} \quad .$$

The Born radii R_i are calculated using a fast and accurate tree-code based approach[77], the surface tension $\gamma = 0.00542\frac{kcal/mol}{\text{Å}^2}$ and the two dielectric constants $\epsilon_p = 1$ and $\epsilon_w = 80$ inside the protein or within the water, respectively.

3.2.2. Bonded Interactions

Bonded interactions are mainly used to describe covalent bond properties like bond distance, angle bending or dihedral rotations. Variables with a superscript 0 denote the value of the asso-

ciated observable within its equilibrium state.

Bond Potential

The bond length fluctuation of two covalent bond atoms i and j can be represented by a harmonic potential:

$$U_{Bond}(\vec{x}) = \frac{1}{2} k_{ij} \left(r_{ij} - r_{ij}^0 \right) \quad . \tag{3.4}$$

Angle Potential

Also the bond angle θ_{ijk} between three bonded atoms i, j and k is described within a harmonic potential:

$$U_{Angle}(\vec{x}) = \frac{1}{2} k_{ijk} \left(\theta_{ijk} - \theta_{ijk}^0 \right) \quad . \tag{3.5}$$

Dihedral Potential

Typical protein force fields also include many dihedral constraints. Improper dihedrals are used to ensure the planarity of planar groups (e.g. aromatic rings) or to favor a structure over its mirror image. Mainly they are described by a harmonic function:

$$U_{iDihedral}(\phi_{ijkl}) = \frac{1}{2} k_{ijkl} (\phi_{ijkl} - \phi_{ijkl}^0)^2 \quad . \tag{3.6}$$

For proper dihedrals, a periodic potential function is used:

$$U_{pDihedral}(\phi_{ijkl}) = k_{ijkl}(1 + \cos(n\phi_{ijkl} - \phi^n)) \quad . \tag{3.7}$$

Proper dihedral potentials are periodic functions, used to either prefer *cis*- or *trans*- configuration for an improved description of hydrocarbon molecules. The multiplicity and the corresponding phase are given by $n \in \mathbb{N}$ and ϕ^n.

Morse Potential

The Morse potential is an anharmonic bond potential that differs from the simple harmonic potential as it has an asymptotic potential well for small distances and a zero derivative for infinite distances [134]:

$$U_{Morse}(\vec{x}) = D_{ij} \left[1 - e^{-\beta_{ij}(r_{ij} - b_{ij})} \right] \quad . \tag{3.8}$$

The bond dissociation energy is given by D_{ij}, the equilibrium distance lies at b_{ij} and the slope of the potential is controlled by the value of β_{ij}.

3.2.3. Restraints

Special restrains are often used to restrict parts of the simulated systems in their attitude or to include knowledge from experimental data. Simple restraints are position or distance restrains, but also more complex restrains are currently implemented in SIMONA like RMSD or secondary structure restrains. All restraints follow the same definition:

- they feature one **collective variable** (CV): $c(\vec{x})$,
- which is restraint by a **distribution**: $D(r)$
- and their contribution to the potential energy is given by a **function**: $U_R(\vec{x}) = D(c(\vec{x}))$.

3.3. Implementation

SIMONA was designed with the idea of being user friendly and at the same time applicable to a variety of problems at once. We could realize this idea by splitting the program up into two different layers. This two-stage concept includes a Python based preprocessor and a powerful C++ kernel. Both, the preprocessor and the kernel are platform independent and can be run on Linux, Windows and OS X.

3.3.1. Preprocessor

The preprocessor is implemented as GUI application that guides the user through the whole XML creation process, making it easy to choose the right force field parametrization, select adequate degrees of freedom and find a proper simulation scheme. Behind the user interface, a Python based preprocessor analyses the input information, creates abstract coordinates, assigns force field parameters, detects degrees of freedom, and finally sets up the whole simulation hierarchy. The preprocessor itself is split into tree distinct stages:

- **Structure**
  In the first stage the structural input data supplied by the user is parsed, based on known residue types. Currently the preprocessor supports the 20 natural amino acids and DNA residues. User specific residues such as pentacene molecules can be supplied easily. The preprocessor sets up an internal object hierarchy containing all structural information like chains, residues or atoms. For each atom, the binding partners are verified and a list of all parameters (e.g. partial charge, Born radii, ...) is stored. By default, these parameters were taken from the PFF specifications, but using a GROMACS topology file, also further specifications (e.g. AMBER) can be used.
- **Moves**
  Having setup up the complete bond structure, the move detection step can be initiated. The preprocessor is able to detect bond stretches, angle rotations and dihedral rotations as well as rigid translations and rotations. But also more complex move classes like cluster

moves or loop rebuilding moves are currently implemented. The assembled move list is then used as a transition function for a Metropolis acceptance criterion.

- **Force field**
  In the final step the force field is created. Besides complete force fields, e.g. PFF0x or AMBER99SB, the user is also able to manually define a list of potential terms. For better operability, each force field term stores its parameters in a distinct object.

Finally, the preprocessor serializes the recently generated data into a XML layer. The resulting data is stored in a human readable XML file which serves as input for the C++ kernel.

3.3.2. XML Abstraction Layer

Input files for the SIMONA kernel are written in the XML language what makes them clearly and user editable (Figure 3.3). A functional input file comprises at least four distinct sections (Figure 3.2):

- **Configuration**:
  The *configuration* section contains the coordinates of all atoms. The C++ kernel only knows non-interacting mass points. Their physical properties are defined in subsequent sections.

- **Moves**:
  The *moves* section holds information about the degrees of freedom of the system (e.g. bonds, angles, dihedrals or rigid body transformations), which can be perturbed during the simulation. The composition of available moves indirectly defines molecular properties, like the system's bond hierarchy.

- **Force field**:
  The *force field* section contains the definition of the potential energy function, including all required parameters. Adding XML elements allows the rapid prototyping of new energy functions.

- **Algorithm**:
  The *algorithm* section encodes the simulation procedure, providing the control flow of the SIMONA kernel. In addition to standard algorithm schemes like Metropolis Monte Carlo or simulated annealing, the XML language also allows the scripting of user defined pathways (Figure 3.4). In this way, new simulation methods can be developed without the need of changes in the C++ kernel.

3.4. Parallelization Strategies

The costs of evaluating non-bonded potential terms grow with the order of $\mathcal{O}(N^2)$. Modern tree code implementations like the Barnes-Hut [135] method can help to reduce these costs to $\mathcal{O}(N \log(N))$ using spatial decomposition. Besides the development of novel and more efficient

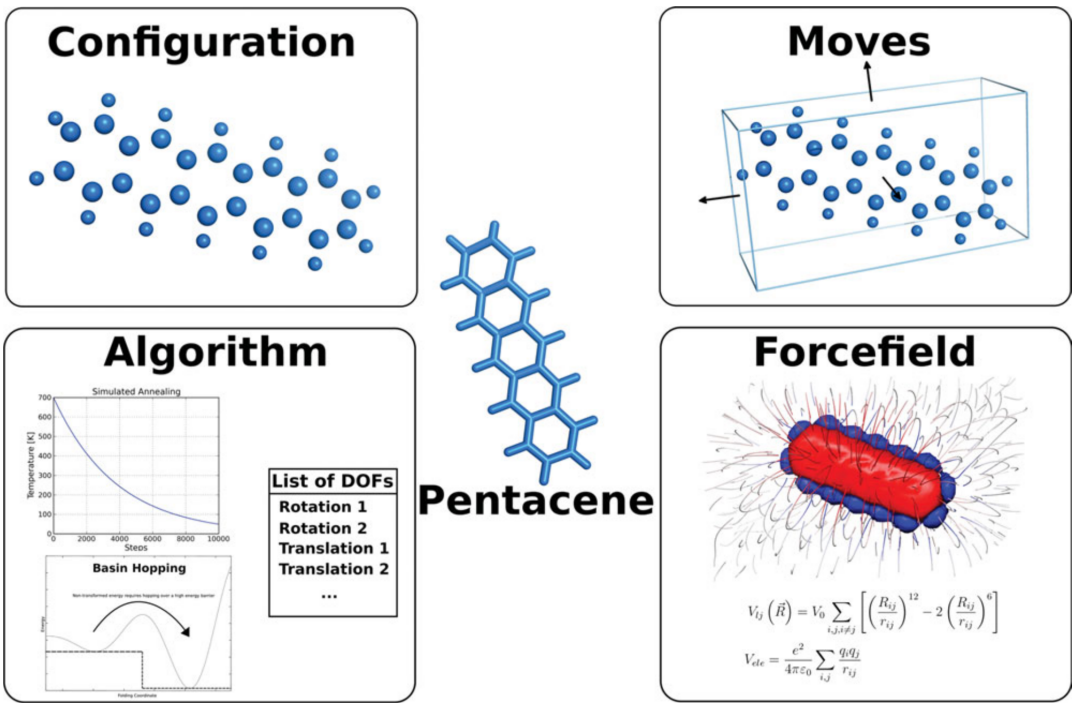


Figure 3.2.: *A valid SIMONA input file consist of four distinct sections. The coordinates of all simulated particles are stored in the* ***configuration****; the available transformations are stored in the* ***moves****; the simulation protocol is defined in* ***algorithm****; the* ***force field*** *part holds the potential function including all parameters. (Copyright © (2012), Wiley Periodicals, Inc.[2])*

algorithms, the parallelization of existing methods also decreases the simulation duration. In SIMONA, we currently have implemented three different approaches:

- **GPU acceleration** via OpenCL: The use of modern hybrid architectures allows the program to use up to 2^{11} threads of modern GPU coprocessors at once. However, these SIMT (Single Instruction, Multiple Threads) architectures have strong limitations regarding branching and memory access.

- **Message Passing Interface** (MPI): MPI is used to spread a program over multiple discrete computing nodes, allowing the use of cluster computers. Actual supercomputers consist of millions of CPUs; the currently fastest one, Tianhe-2 (3,120,000 CPUs), strikes with a theoretical peak performance of $\approx 55\text{PFlop/s}$[136]. The inter-node communication is realized over fast interconnects (TH Express, Infiniband or custom approaches).

- **OpenMP**: With OpenMP, a program can simply unroll loops into multiple threads on the local compute node. It requires thread safe programming and handles the inter-process communication via shared memory.

Modern supercomputers allow the combination of MPI with OpenMP for the optimal allocation of compute nodes. This makes it possible to allocate hundred-thousands of CPUs in one program.

```
<simona_input>
  <simona_configuration>
    <coord residue_name="MET" residue_id="0" chain_id="0" name="N" id="0">
      <X>1.18</X>
      <Y>-10.03</Y>
      <Z>-3.49</Z>
    </coord>
  </simona_configuration>
  <simona_moves>
    <simona_dihedrals>
      <DihedralSpec>
        <unique_id>0</unique_id>
        <DihedralInfo>
          <predecessor>4</predecessor>
          <bond_start>6</bond_start>
          <bond_end>9</bond_end>
          <successor>12</successor>
        </DihedralInfo>
        <AtomSet>[...]</AtomSet>
        <type>41.MET.chi2</type>
      </DihedralSpec>
    </simona_dihedrals>
  </simona_moves>
  <energy_models>
    <forcefield id="0" name="gromacs">
      <Coulomb>[...]</Coulomb>
      <GBSolvationStill>[...]</GBSolvationStill>
      <PowerBorn>[...]</PowerBorn>
      <NPSasaEnergy>[...]</NPSasaEnergy>
      <LennardJones>[...]</LennardJones>
      <PowerSasa>[...]</PowerSasa>
      <DihedralPotential>[...]</DihedralPotential>
    </forcefield>
  </energy_models>
  <Algorithm>
    <RepeatedMove>
      <tscaling>geometric</tscaling>
      <repeats>10000</repeats>
      <tstart>300.0</tstart>
      <tend>300.0</tend>
      <TransformationSequence repeats="1" weight="1.0">[...]</
          TransformationSequence>
    </RepeatedMove>
  </Algorithm>
</simona_input>
```

Figure 3.3.: *The image shows the translation of the four distinct section of a basic SIMONA input into a XML abstraction layer. The file was strongly stripped to enhance readability. The four distinct sections of a XML file are highlighted.*

```
<Algorithm>                                          for i in 0..12000 do
  <RepeatedMove>
    <repeats>12000</repeats>
    <tstart>450.0</tstart> <tend>150.0</tend>          #geometrical temperature scaling
    <tscaling>geometric</tscaling>                     temperature = temperature_before*temp_factor
    <TransformationSequence>                           Do all transformations in a row
      <ConditionalTransformation>                        Create a copy of the current configuration
                                                         Apply following transformations to copy
        <TransformationChoice>                             Choose one transformation of
          <SetTranslationRandom wt="1.0">                    A random translation (with a weight of 1)
          <SetRotationRandom wt="1.0">                       A random rotation (with a weight of 1)
          <TransformationChoice wt="20.0">                   Choose one (dihedral perturbation) (weight 20)
            <SetDihedralRelativeRandom wt="1.0">                 A dihedral perturbation (weight 1)
            <SetDihedralRelativeRandom wt="1.0">                 A dihedral perturbation (weight 1)
            <SetDihedralRelativeRandom wt="1.0">                 A dihedral perturbation (weight 1)
              ....                                               ... (complete List of dihedral angles)
          </TransformationChoice>
        </TransformationChoice>
       <MetropolisAcceptanceCriterion>                 Evaluate energy difference between copy and original
          <energymodel_nr>0</energymodel_nr>           Accept or reject based on Metropolis criterion
          <kB>0.0019858775</kB>
       </MetropolisAcceptanceCriterion>
      </ConditionalTransformation>
      <TransformationSequence>                         Do all transformations in a row
        <EnergyOutput>                                   Output the energy
        <ConfigurationOutput>                            Output the snapshot for a trajectory
      </TransformationSequence>
    </TransformationSequence>
  </RepeatedMove>
</Algorithm>
```

Figure 3.4.: *A closer look to the algorithm section demonstrates the potentiality of generating new simulation protocols. A default Monte Carlo path is shown on the left, translated into pseudo code on the right. In each step either a rigid rotation or translation or a random dihedral rotation is performed. The proposed structure will be accepted according to the MH criterion at a certain temperature that scales geometrically from* $450K$ *to* $150K$ *over* $12,000$ *steps. (Copyright © (2012), Wiley Periodicals, Inc.[2])*

3.4.1. GPU Acceleration

In a study for demonstrating the benefits gained by the use of GPU accelerators for the simulation of proteinous systems, we compared two graphics boards of the two leading manufacturers with a twelve-core CPU system[3]. All simulations were carried out using the same code path in an OpenCl implementation (Figure 3.7), where N denotes the number of atoms and W_S stands for the workgroup size:

1. The N workitems were grouped into $\frac{N}{W_S}$ workgroups.
2. In an *ordered* manner, all N workitems load their assigned coordinate into their register.
3. For $i = 1, ..., \frac{N}{W_S}$ all workgroups perform the same loop:
   a) Each workgroup loads segment i *ordered* into its local cache.
   b) The workitems use broadcast reads to load the segments coordinates from local cache.
   c) Each workitem calculates the interaction energy for its assigned particle with all other particles in the workgroup.
4. After all segments have been evaluated, the sum over all energies will be returned.

The performance of the three systems strongly depends on the number of simulated atoms and the optimal choice of the workgroup size for each system (Figure 3.5). We carried out multi-

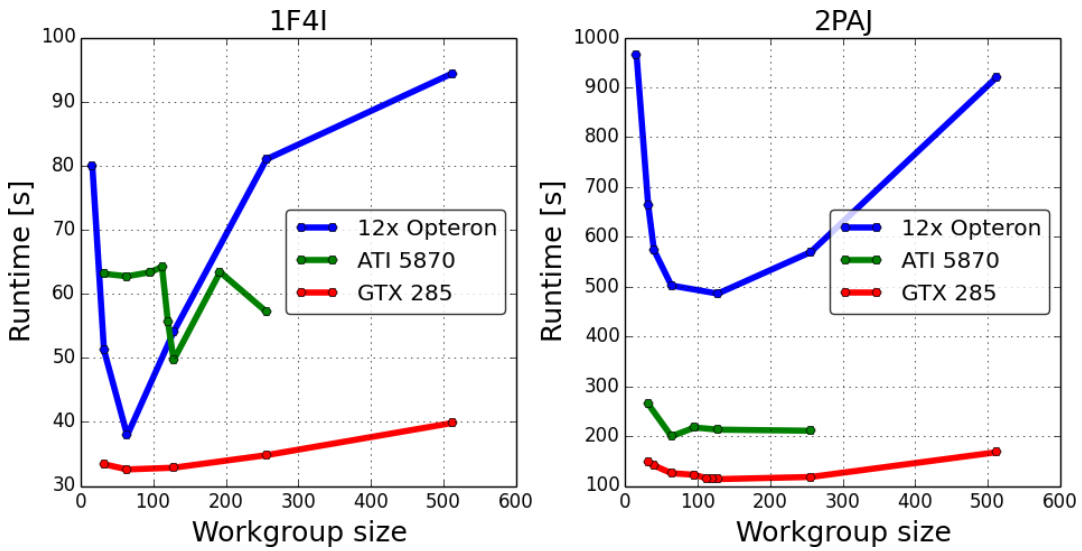


Figure 3.5.: *The runtime of the OpenCL accelerated code shows a strong dependency on the workgroup size for CPU systems and a smaller one for GPU systems. Where in the smaller protein system 1F4I (365 atoms), the twelve core CPU system still can beat the ATI5870 GPU, it has to surrender for larger systems like 2PAJ (3194 atoms).*

ple simulations for different sized proteins within an own implementation of the SMOG force field[137]. The parallelization approach was applied to the non-bonded Lennard Jones potential term, the rate limiting step of the force field. For larger protein systems we could achieve speedups of up to ≈ 150 for the GPU coprocessor in comparison with the CPU-only system (Figure 3.6).

3.4.2. Message Passing Interface

SIMONA mainly uses the MPI layer for the parallelization of simulation methods. Instead of parallelizing the energy evaluation, we developed several methods to enhance the efficiency of the default Monte Carlo simulation scheme. Multiple Try Metropolis directly interferes with the Markov core by generating multiple proposals in each step instead of a single one (Section 5.2.2). A parallel tempering approach lets the user simulate a set of temperatures at once, allowing the exchange of structures between two populations what lowers the overall barrier heights and increases the phase space exploration speed (Section 6.2.5). The schematic MPI implementations of these two approaches are shown in Figure 3.8 and 3.9.

3.4.3. OpenMP

Other than MPI, OpenMP is a shared memory approach to parallelize the execution of code blocks via threadding [138, 139]. In SIMONA, OpenMP is used to spread the evaluation of pair-interaction terms on available CPUs what significantly reduces the energy evaluation times for non-bonded energy functions (Figure 3.10). However, the benefits of such a parallelization are strongly limited by the number of local processors. Currently we are using four processors for a single energy evaluation of a small protein. The OpenMP approach can be easily combined

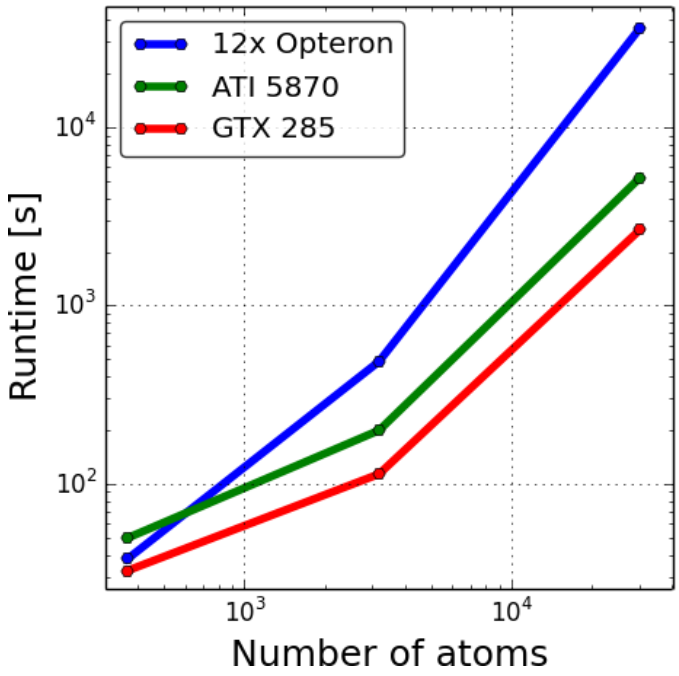


Figure 3.6.: *The double exponential graph shows the runtime of a 10,000 steps MC simulation depending on the number of atoms and the used platform. Each points was taken from the optimal choice of the workgroup size. The measuring point for the largest system 2VZ8 on CPUs was extrapolated based on a 200 steps simulation. With increasing number of atoms, the CPU system more and more legs behind both GPU systems.*

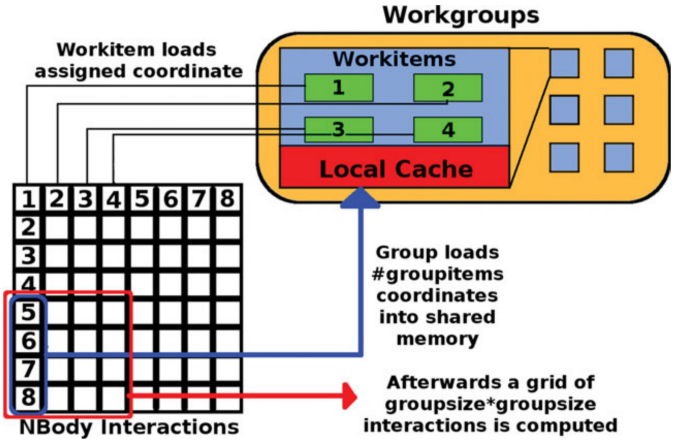


Figure 3.7.: *The N×N contributions of the non-bonded Lennard Jones potential are broken into blocks of which each one is evaluated simultaneously by one workgroup within its local memory. Each core of the workgroup is associated with one atom and calculates its interactions with all other atoms of the workgroup. This process is repeated until each block is evaluated. (Copyright © (2012), Wiley Periodicals, Inc.[2])*

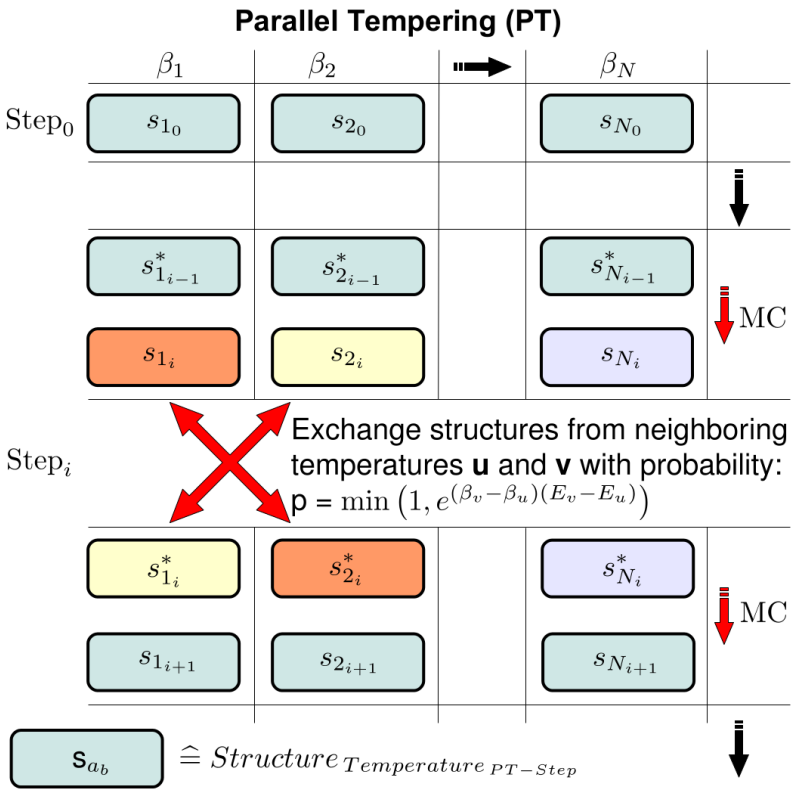


Figure 3.8.: *In the parallel tempering simulation scheme, multiple MH simulations are performed simultaneously at different temperatures on separate processors. After a certain step, adjacent temperatures exchange their configuration proportional to a modified MH acceptance criterion. However, to spare bandwidth the MPI implementation exchanges temperatures instead of configurations. The temperature range should include high temperatures to cross every barrier as well as low temperatures to freeze in local minima. For best PT exchange rates the temperatures should be distributed with an exponentially growing distance.*

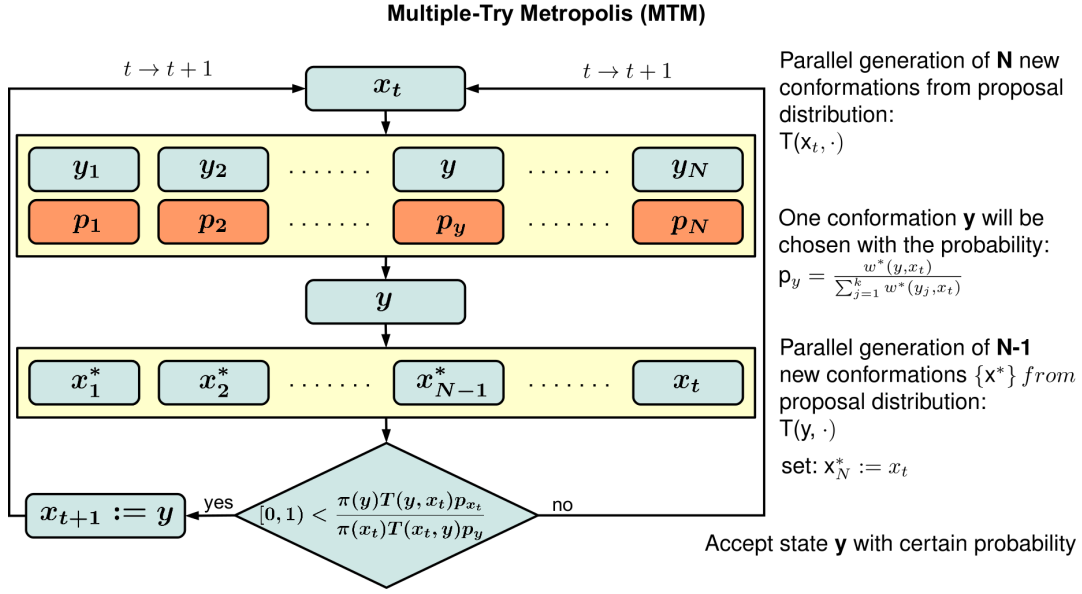
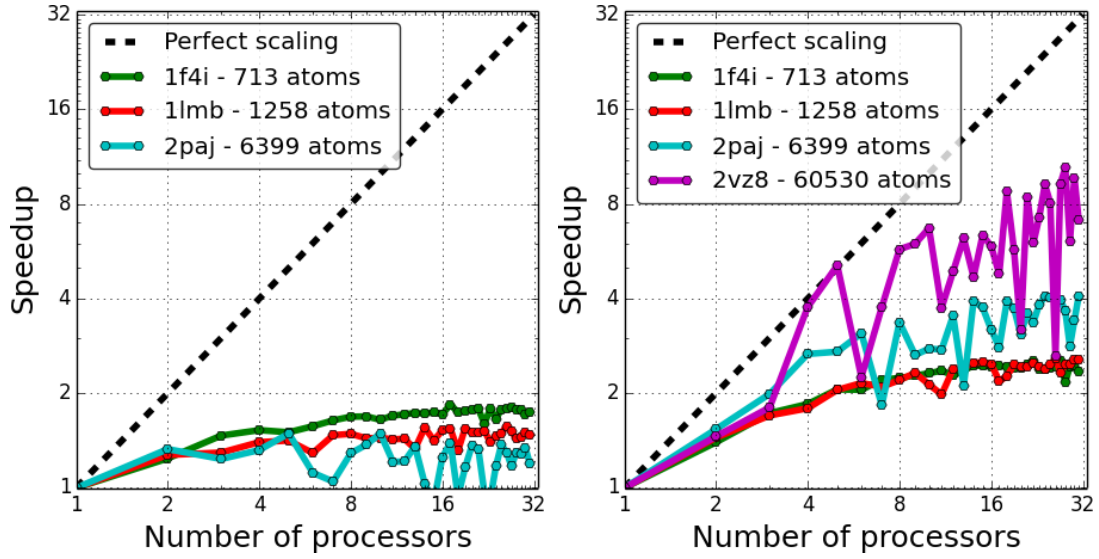


Figure 3.9.: *The Multiple Try Metropolis method comprises two proposal steps. In the first step, N new conformations are generated of these one is selected for further investigation. Based on the picked trial, $N-1$ new states are then generated in the second step. Finally, the trial is accepted proportional to a certain acceptance criterion based on the energy distributions of both steps. In each proposal step, the generation of the trials can be spread over N processors allowing a parallel energy evaluation.*



(a) *Version 1* **(b)** *Version 2*

Figure 3.10.: *Depending on the size of the simulated protein, the OpenMP parallelization approach leads to a performance boost an with adequate efficiency. For larger protein systems, the benefit from parallelizing a combined potential term including all non-bonded interactions (3.10b) instead of a parallelization of only the Lennard Jones term (3.10a) clarifies. Our current simulations of the Villin headpiece run on four processors.*

with an existing MPI-based simulation structure to be perfectly fitted for the use on modern supercomputers.

3.5. Conclusions

The simulation toolkit SIMONA, presented in this chapter, implements various simulation techniques within numerous popular force fields[2]. It was designed with the intent to craft a highly adaptive framework, by simultaneously being as user friendly as possible. On the one hand, its flexibility is given by the XML based input structure of SIMONA that allows computational versed users to develop new simulation approaches by scripting the XML layer and on the other hand the graphical user interface (GUI) represents a user friendly and clickable preprocessor to set up individual simulations without prior knowledge of the simulation package.
The C++ code base of SIMONA was optimized for modern parallel CPU architectures and their instruction sets, as well as GPU coprocessors and runs on a diversity of operating systems: Linux, Windows, OS X (Proceeding 1, 2). It was published under an academic license and can be downloaded for free at[1]:

`http://www.int.kit.edu/nanosim/simona.`

All Monte Carlo based simulation techniques presented in this work were implemented in the SIMONA package. In further investigations (Chapter 6) we could demonstrate that the Monte Carlo based simulation approach successfully overcomes the time scale problem of the more

[1] Statistics about the source code can be found in the Appendix A.1

popular molecular dynamic method. Using SIMONA, we could successfully perform a complete thermodynamic characterization of a ultrafast folding protein within an atomistic, physical force field.
Additional to common MH approaches, we could extend massively parallel simulation schemes to be applicable to the most complex problem of protein folding (Chapter 5). Using SIMONAs MPI capabilities, we could utilize this method to simulate multiple reversible folding and unfolding transitions faster than ever before.

4. Structure of the Twin Arginine Translocase

4.1. Motivation

The translocation of proteins across cytoplasmic membranes is critical for the correct function of living cells. In addition to the general Sec protein-translocation pathway[1], a second pathway, the twin arginine translocase (Tat) has been discovered recently[140, 141]. Tat translocases comprise two protein types, a hexahelical TatC-type and members of the TatA family (TatA and TatB). In spite of the large similarity of the TatA and TatB proteins, their tasks within the Tat pathway differ[142]. Apart from the functional dissimilarity, the expression level of TatA is much higher compared to the associates, TatB and TatC[143]. Homologous proteins can be found in many archaea, bacteria, plant chloroplasts and mitochondria[144, 145]. In contrast to the Sec system, which transports unstructured proteins driven by ATP hydrolysis in combination with the transmembrane electrochemical gradient, Tat translocases exclusively use the ΔpH-gradient to carry fully folded proteins across membranes [146–149].
The importance of transporting folded proteins across membranes can be illustrated by means of enteric bacteria, where the energy metabolism strongly depends on the electron transfer across the cytoplasmic membrane. Being redox-active and thus able to perform their catalytic or electron transfer functionality, proteins have to bind cofactors like nucleotides or metal ions. In most cases this requires the protein to fold and makes it unsuitable for the Sec system transport[150]. However, this requires the Tat system to generate much larger pores.
The assembly process of TatA pores is triggered by a signal peptide that binds on one end to the cargo protein and is recognized by the TatC molecule through a *twin arginine* pattern on the other end. Afterwards a dynamically sized pore consisting of multiple TatA subunits spontaneously forms around the cargo. After the completion of the translocation of the cargo through the membrane, the TatA pore dissolves without harming the membrane.
Due to the high heterogeneity of the formed pores, the experimental resolution of structural details of the Tat pore complex is limited to single TatA monomers or fragments thereof. Except for cryo-EM data which pointed out that a single pore appears as homo-oligomeric TatA complexes[151], only the TatA monomer could be analyzed. It comprises an amphiphilic helix (APH), a helical transmembrane segment (TMS) and a densely charged, unstructured region (DCR)[152].

In collaboration with the experimental workgroup Ulrich at the KIT and the theoretical MD group Ruggerone at the Università di Cagliari we investigated the sequential constitution of the TatA monomers and noticed an intrinsic complementary charge pattern spread over all TatA homologs (Figure 4.1)[4]. Although most of them occur in transmembrane regions, their sequences comprise 40% of charged residues in cytoplasmic domains instead of the statistically expected 22.5%. However, they are nearly uncharged in total. In the examined organism, bacillus sub-

name	sequence of charges along the APH [and TMS]	sequence of charges along the DCR [and C-terminus]	K/R	D/E	net
B.subtilis_TatAd	+.-.+.+.-.+.+.	- - + - - +.-[.+.- +]	7/2	2/6	+1
B.subtilis_TatAy	+ +.-.+.-.+ -.+.+.	- - -.- - + + - -	8/1	3/6	0
E.coli_TatA	+ +.-.+.+ +.	- - -.+.- +.-[-.+.- +.-.-.+.- -.+ +.- + -]	12/1	10/4	-1
E.coli_TatE	[-.+]+ +.+.-.+.+ +.	- - -.+ +-.-.- +[.+ -]	10/1	6/3	+2
V.cholerae_TatA	+ +.+.-.+.+ +.	- - -.+ -.-.-[.+.-.+.-.+ + - + -]	11/1	4/7	+1
Y.pestis_TatA	+.+.-.+.+ +.	- -.- +.-.-[.+.- +.+.- -.+.- + -]	10/1	6/5	0
S.enterica_TatA	+ +.-.+.+ +.	- - -.+.- +.-[.-.+.- +.-.+ + - -.+.- + -]	13/0	10/3	0
H.pylorii_TatA	+ +.-.+.+.+ +.+	- - - - -.+.-.+.-[.+.-.-.+.+.-]	12/0	3/8	+1
A.tumefaciens_TatA	+.+.-.-.+.+.+ +.	- - -.-.+.-.+.- -[.+]	8/1	6/3	0
P.aeruginosa_TatA	[-.+]+ +.+.-.-.+.+ +.	- - - - + + -.-[.+.- -.+ + -]	9/3	7/5	0
H.influenzae_TatA	+ +.+.-.+.+ +.+	- - - +.+ -.-.+.-[.-.+ +.+.+ + - +.+]	13/4	4/5	+8
M.tuberculosis_TatA	+ +.-.+.+.+.-.+	-.-.+.-.-.+.-.-.-[.+]	5/5	3/6	+1
P.sativum_Tha4	… [-.]+ +.-.+.+.+	-.-.-.+ +-.- [-.-.- +.-.+.+ -]	13/4	1/12	+4
E.coli_TatB	[-.-.]+.+.+.-.-.+.-.-.+ +.- +.-.+.- -.+.-.+ +.-.- +.- -.+…		13/5	8/16	-6
P.sativum_Hcf106	… [-.]+.-.+.+.+ -.+ -.-.+ -.+.- + -.- -.+.-.-.+.- -.+.- -.+…		20/8	7/20	+1

Figure 4.1.: *Charge zipper motifs found in TatA homologs are highlighted. The charge pattern on the APH (blue) is reflected inverse on the DCR (red). Further residues lying between adjacent charges are denoted as dot. copyright © (2013), Elsevier[4].*

tilis, the observed charge pattern forms a ladder consisting of seven charge pairs between the APH and the DCR of $TatA_d$. The successive closure of this zipper like salt-bridge motif would attach both adjacent molecule parts to each other forming a hairpin like structure. Combining multiple $TatA_d$ hairpins through intermolecular salt bridges would then result in an amphiphilic palisade. These palisades, with a height that properly fits the membrane thickness, could be the prerequisite for building pore complexes and lining them up in the membrane.

A charge zipper motif can be discovered as a specific sequence pattern in a protein's primary sequence. For the description of globular protein conformations alternating charge motifs have often been referred to as "charge zipper", "salt bridge zipper", "electrostatic zipper", or "ionic zipper" motifs in the past. Salt bridges were observed to stabilize interfaces between separated subunits like in the pyruvate dehydrogenase multienzyme complex[153, 154], or for dimerization[155, 156]. Also the denaturation of folded proteins may require a successive destruction of formed salt-bridge motifs[157]. However, a direct correlation between the occurrence of charge zippers in the primary sequence of a protein and its corresponding structure was never shown before.

Queries to the Uniprot database[158] revealed that 75% of the entries also feature these *charge zipper motifs*. In our investigations, we demonstrated their importance for membrane associated proteins and protein complexes, especially for the self-assembly process of oligomeric $TatA_d$ translocation pores. In extensive mutation studies, our experimental partners could subsequently identify the binding pattern of the charge zipper motif within the $TatA_d$ pore complex. Especially the use of charge inversion mutations within the proposed oligomerization interface provided a reliable tool for the determination of the exact binding pattern of salt bridges. This technique not only allowed the identification of both residues participating in a salt bridge, but also elucidated, whether the salt bridge is formed intramolecularly or intermolecularly between two adjacent molecules.

We could then add the obtained salt-bridge pattern as a constraint in our MD simulations. For this purpose, we developed an implicit membrane-pore force field, which allowed us to simulate the self-assembly and oligomerization process of dodecameric $TatA_d$ pore complexes.

4.2. Methods

For our investigations of the twin arginine translocase pore forming complex, we combined experimental knowledge with a theoretical description using molecular dynamics simulations. For the validation of the proposed charge patterns, our experimental partners have prepared multiple charge mutations for the $TatA_d$ system followed by gel electrophoresis analysis to determine their influences on the oligomeric stability. Details about the experimental procedures can be found in[4].
The dominant driving force for the structural formation of membrane proteins are the electrostatic interactions and the interaction with the environment. However, the theoretical description holds some complications. Explicit all-atom modeling of the electrostatic environment is strongly limited due to the huge number of participating atoms; but also an implicit description of the environment based on GB-methods barely reaches time scales on which the fastest transitions, like folding of transmembrane helices, occur.
Thus, for the simulation of complex mechanism like the formation of a pore within a membrane we developed a new and fast implicit membrane-pore force field. We describe the protein within an atomistic, structure based potential[159]. The interaction with the environment is modeled by a hydrophobic slab potential acting on the hydrophobicity of each amino acid[160]. Oligomerization is driven by an additional contact-based potential derived from experimental knowledge. The implicit membrane-pore force field comprises three distinct potential terms, each designed to describe a certain physical property of the system:

- **SMOG structure based potential**:
  The SMOG potential is a Go-Potential, which contains multiple terms both to ensure the physical reality of the protein by stabilizing covalent bonds and ensuring Fermi-exclusion and to stabilize the monomer's secondary and tertiary structure using contact information.

- **Implicit membrane-pore potential**:
  This potential approximates the interactions between the simulated proteins and the environment.

- **Salt-bridges constraints**:
  Based on experimental knowledge, additional constraints model the pairwise interaction of salt bridges within protein complexes.

4.2.1. SMOG: Structure-based Potential

Structure based potential are widely used to describe protein folding and lead to correct dynamical information of the folding process[161–164]. Based on a contact map, generated using a

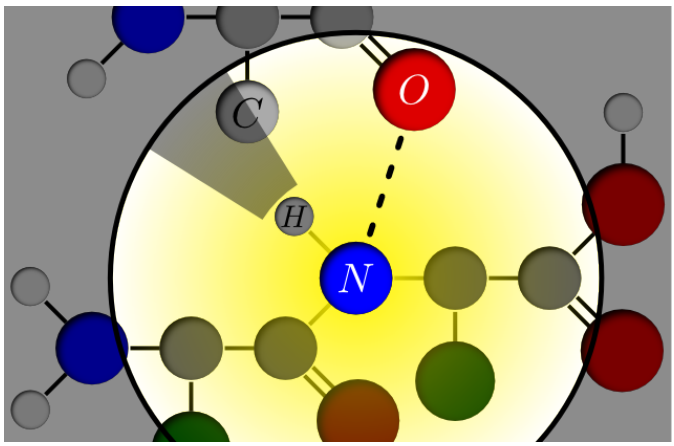


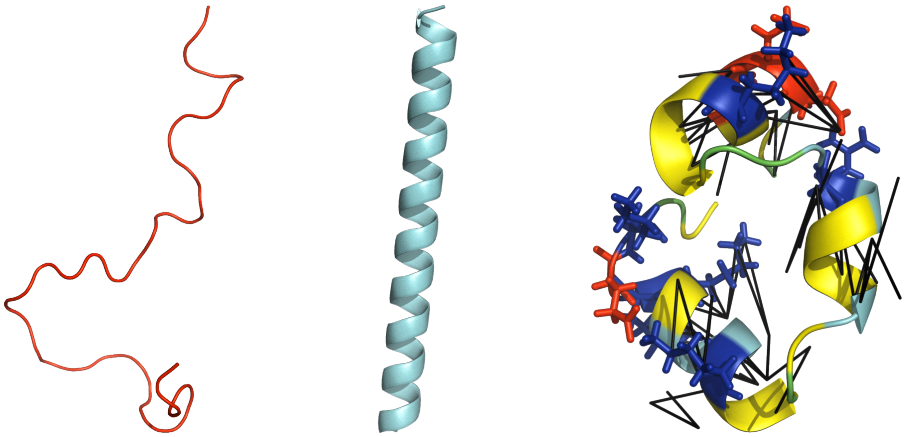
Figure 4.2.: *To create a contact list for atom N, all atoms within a distance of $D_{cut-off} = 6\text{Å}$ are considered as a potential contact, if their sequential residue distance is bigger than 3. To sort out nonphysical contacts, a radial light source is positioned at its center. All atoms are represented as spheres with radius $r_{Atom} = 1\text{Å}$ and can draw shadows onto other atoms ($H \rightarrow C$). Shadowed or partially shadowed atoms will be discarded. In this example only O remains in N's contact list.*

native structure (Figure 4.2), the potential energy of a structure is given by:

$$
\begin{aligned}
V = & \sum_{\text{bonds}} \varepsilon_r (r - r_0)^2 + \sum_{\text{angles}} \varepsilon_\theta (\theta - \theta_0)^2 + \sum_{\text{impropers / planar}} \varepsilon_\chi (\chi - \chi_0)^2 \\
& + \sum_{\text{backbone}} \varepsilon_{BB} F_D(\phi) + \sum_{\text{side chain}} \varepsilon_{SC} F_D(\phi) \\
& + \sum_{\text{contacts}} \varepsilon_C \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r} \right)^{6} \right] + \sum_{\text{non-contacts}} \varepsilon_{NC} \left(\frac{\sigma_{ij}}{r} \right)^{12} \\
F_D(\phi) = & \left[1 - \cos(\phi - \phi_0) \right] + \frac{1}{2} \left[1 - \cos(3(\phi - \phi_0) \right] \qquad ,
\end{aligned}
$$

where all variables with a subscripted 0 denote the native value determined from the experimental structure. The remaining parameters can be found in[159]. Please note that the application of this Hamiltonian results in simulations within a reduced units space, which means that the folding equilibrium temperature lies at $T_{\text{folding}} = \frac{1}{k_B} \approx 110K$.

The limitation of this approach is that the native structure must be known a-priori as the system's Hamiltonian is build using its contact information. Though this hinders one to perform de-novo protein folding simulations, it allows us to describe a protein within a artificial contact map based on other known information. Secondary structure elements like transmembrane helices can be determined experimentally[152] and then be modeled using a default set of secondary structure parameters, such as $\phi - \psi$ pairs. The resulting contact map for a TatA_d monomer, having two helical regions, can be generated without any prior knowledge except its amino acid sequence (Figure 4.3). For the simulation of oligomeric complexes, this contact map can be applied to every single monomer, allowing them all to fold individually and simultaneously assume their secondary structure elements. Except for Van der Waals interaction, this potential does not include any intermolecular interaction.



(a) *Coiled residue chain.*

(b) *Applied secondary structure.*

(c) *Resulting constraint potential.*

Figure 4.3.: *SMOG: An exemplary force field setup for the Villin headpiece invokes the generation of a random initial structure based only on the protein's primary sequence (4.3a). Similar to the simulations in the following section, a single helix conformation was generated using secondary structure distance constraints (4.3b). Region crossing constraints had to be pruned from the resulting SMOG potential. All remaining contacts are shown in (4.3c).*

4.2.2. Implicit Membrane / Pore Potential

For the simulation of membrane associated protein complexes, we incorporate interactions between the proteins and the membrane. To save computational costs and allow the investigation of large complexes as well as long time scales, we chose an implicit description of the membrane. We approximate all atomic interactions between the membrane and an amino acid by a linear function along the membrane normal scaled with the hydrophobicity coefficient[160] of the particular residue. Parallel to the membrane normal, the potential function has the shape of a double well potential. Further, we implemented the incorporation of a cylindric pore spanning the membrane. We modeled this interaction by using a second radial double well potential in respect to the pore. This separates the simulation into one (hydrophobic) membrane lying in between two (hydrophilic) water regions, which are connected through a (hydrophilic) cylindric pore piercing the membrane layer (Figure 4.4). In a three-dimensional notation the resulting potential energy is given by:

$$
\begin{aligned}
V_{\text{membrane}} &= F_{\text{res}} \cdot V_r \cdot V_z \\
V_r &= 0.5 \left(\tanh \left(\frac{r - r_{\text{width}}}{r_{\text{slope}}} \right) - \tanh \left(\frac{r + r_{\text{width}}}{r_{\text{slope}}} \right) \right) + 1 \\
V_z &= 0.5 \left(\tanh \left(\frac{z - z_{\text{width}}}{z_{\text{slope}}} \right) - \tanh \left(\frac{z + z_{\text{width}}}{z_{\text{slope}}} \right) \right) + 1 \qquad ,
\end{aligned}
$$

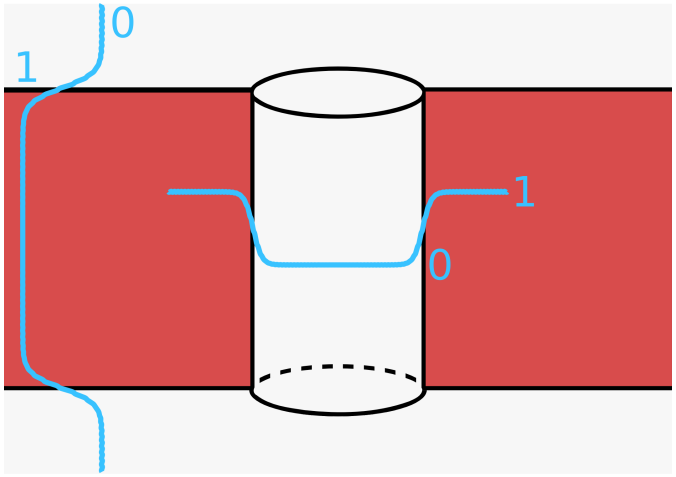


Figure 4.4.: *A three-dimensional membrane-pore potential is assembled by multiplying two linearly independent potential well functions. The first well lies in the direction perpendicular to the membrane surface, whereas the second well lies in the plane spawned by the pores major axis as normal vector. The resulting potential divides the space into a membrane region (red $\equiv$ 1) and water / cargo regions (gray $\equiv$ 0), with smooth intersections in between. Multiplying this function with the hydrophobicity coefficient of an amino acids results in a suitable approximation for a membrane-pore potential.*

and the resulting force that acts on the center of mass of each amino acid derives to:

$$\begin{aligned}
\vec{F} &= F_r \cdot \vec{e}_r + F_z \cdot \vec{e}_z \\
F_r &= -V_z \cdot \frac{F_{\text{res}}}{r_{\text{slope}}} \cdot 0.5 \left(\text{sech}^2 \left(\frac{r - r_{\text{width}}}{r_{\text{slope}}} \right) - \text{sech}^2 \left(\frac{r + r_{\text{width}}}{r_{\text{slope}}} \right) \right) \\
F_z &= -V_r \cdot \frac{F_{\text{res}}}{z_{\text{slope}}} \cdot 0.5 \left(\text{sech}^2 \left(\frac{z - z_{\text{width}}}{z_{\text{slope}}} \right) - \text{sech}^2 \left(\frac{z + z_{\text{width}}}{z_{\text{slope}}} \right) \right) \quad .
\end{aligned}$$

The two parameters *r* and *z*, each occurring with a subscripted *width* or *slope*, define the shape of the potential well of both the membrane and the pore. Simulations of the described force field would now allow to simulate the association of single monomers to either the membrane surface or the radial pore.

4.2.3. Salt-Bridge Constraints

To allow for the simulation of oligomeric system, the force field described in the section above must be complemented by intermolecular force field terms. Based on the postulated *charge zipper motifs* one additional constraint is added for every single salt bridge, modeling the interaction between monomers and the arrangement of secondary structure elements within a single monomer. To allow simulations of large conformational changes, e.g. initially positioning the monomers spatially far from each other, we define a salt bridge by a long range Morse potential:

$$V_{\text{salt bridge}} = D_e \left(1 - e^{-\beta(R - R_e)} \right) \quad , \tag{4.1}$$

with $D_e = 30 \frac{kJ}{mol}$; $R_e = 2.353\text{Å}$ and $\beta = 0.1\text{Å}^{-1}$. These parameters were chosen soft enough to not induce unphysical results by destroying secondary structure information but sufficiently

```
POSITION LIST <g1> LINE_DIST XYZ 9.923503 10.124253 9.266094 9.923503
    10.124253 12.266094
POSITION LIST <g1> LINE_POS XYZ 9.923503 10.124253 10.766094 9.923503
    10.124253 13.766094
g1->
1 2 3 4 5 6 7 8 9 10 11
g1<-
LWALL CV 1 LIMIT 0.0 KAPPA 15.000000 EXP 2 EPS 0.300000 OFF 4.500000
LWALL CV 2 LIMIT 1.0 KAPPA 15.000000 EXP 1 EPS 0.300000 OFF 1.500000
```

Figure 4.5.: *Plumed example code parametrizing a residue described by the center of mass of a group of atoms* $1-11$ *within the implicit membrane-pore potential. Modified collective variables allow to describe the residue within a three-dimensional membrane-pore potential shown in figure 4.4.*

strong to bind adjacent monomers. The complete force field then describes the intra- and inter-monomeric interactions as well as the interaction between the proteins and the environment. At the same time, the proteins mostly retain their secondary structure configuration.

4.2.4. Implementation

The SMOG potential is currently available for the GROMACS MD simulation toolkit. The parametrized potential can be modified by rewriting GROMACS topology files, which contain the whole potential function[107, 137]. Additional salt-bridge potentials are introduced by extending this topology file. The complex membrane-pore potential described in section 4.2.2 was implemented using the metadynamics framework plumed, which is most commonly used for improved sampling techniques on biological systems[165]. It is available as a patch for the GROMACS package and grants access to complex functions based on collective variables (CV). A CV is equivalent to constraint functions implemented in SIMONA (Section 3.2.3), defined by a projection function onto a one dimensional scale and a distribution to contribute to the potential energy of the system.
To create our three-dimensional membrane-pore potential, we use two different CVs: *LINE_POS* and *LINE_DIST*. Further we extend the single-well *LWALL* distribution to reflect a double sided potential well and add the possibility to multiply sequential CVs. This ability to multiply sequential CVs is essential for creating the three-dimensional membrane pore environment (Figure 4.4). A sample input for a residue within the membrane potential also including a pore is shown in figure 4.5.

4.2.5. Simulation Protocol

In order to perform an implicit membrane simulation using charge zipper contacts, we setup a molecule specific force field. As the structure based part of the force field depends on a native contact map, we create an initial protein structure. For each known secondary structure region, we apply main-chain dihedral angle values according to the following list:

- **α-helix**: $\phi = -48.0°$ and $\psi = -57.0°$

- **3, 10-helix**: $\phi = -49.0°$ and $\psi = -26.0°$
- **β-sheet**: $\phi = -135.0°$ and $\psi = 135.0°$
- **coil**: $\phi = -180.0°$ and $\psi = 180.0°$.

After generating the contact map for the molecule, we clean it by removing all contacts affecting either coiled regions or atom constraints related to conflicting secondary structure motifs. The removal of these contacts is essential for the creation of the structure based potential to not accidentally introduce wrong interactions. The implicit membrane potential, optionally comprising a cylindric pore, can be constructed by using the sequential information of the molecule and adding one CV for each residue according to section 4.2.2. We introduce the charge zipper motif by adding one CV per proposed contact as explained in section 4.2.3. For multimeric complexes, we map the intra-molecular contacts to all neighboring monomer pairs.

We then performed simulated annealing MD simulations, cooling down from $50K$ to $10K$ within $10ps$. Please note that both, temperature and time scale are given in reduced units. The structure based potential introduces reduced units, which results in a folding / unfolding equilibrium at $T_{folding} \approx 110K$. At a temperature of $T_{assemble} \approx 50K$ we could observe both membrane alignment and assembly of multimeric structures by simultaneously conserving the secondary structure of the monomers.

4.3. Results

We will show that electrostatically driven processes within protein-membrane systems like membrane alignment, self-assembly or the formation of transmembrane pore complexes can be described based on the observed charge zipper motifs. Including the intrinsic, alternating charge pattern, which is encoded in the primary sequence of single TatA monomers into our simulations, we could successfully describe the pore-formation process of dodecameric $TatA_d$ pores.
A search for charge zipper motifs within the Uniprot database[158] that contains over 533.000 sequences, yielded 400.000 mostly redundant protein sequences enclosing a potential match for a charge zipper motif comprising at least five complementary charge pairs. To illustrate the importance and the reference of charge zipper motifs for the structural assembly of membrane proteins, we selected a few interesting biological systems. Using these systems, we elucidate the membrane alignment behavior of smaller membrane proteins including the formation of smaller pores and demonstrate the transferability of the presented implicit membrane force field:

- Twin arginine translocase: **$TatA_d$ pore formation** (Section 4.3.1)
- The bacterial stress-response peptide: **TisB** (Section 4.3.2)
- The anionic antimicrobial peptide: **Dermcidin** (Section 4.3.3)
- The structural glycoprotein of pestiviruses: **E^{rns}** (Section 4.3.4)

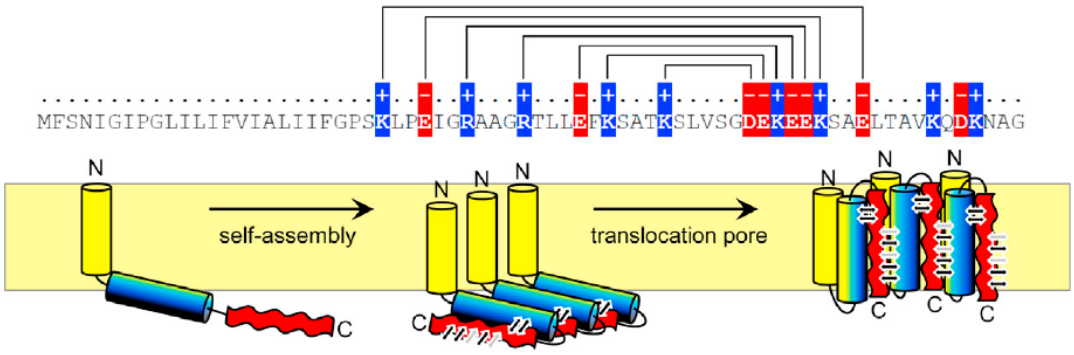


Figure 4.6.: *TatA: charge pattern. Row 1: K41, K45; row 2: K25, E28, R31, R35, E39; separated by bulky L38. Copyright © (2013), Elsevier[4].*

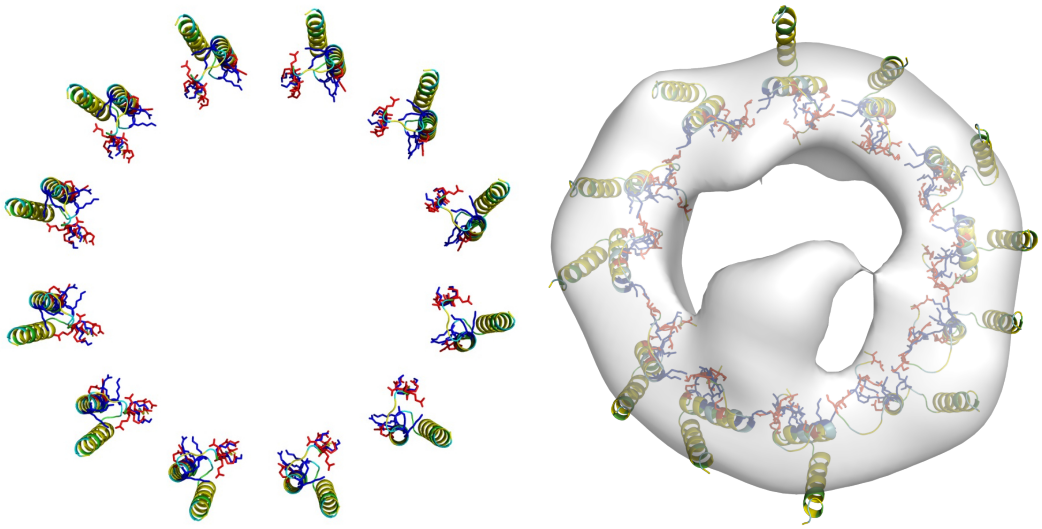
4.3.1. TatA Pore Formation

By examining the TatA sequence, we could observe a charge zipper motif with complementary charges on the amphiphilic helix (APH) and the densely charged, unstructured region (DCR). The charge pattern suggests a network consisting of seven salt bridges. The discrepancy in the number of residues between both opposed charge motifs, more than 21 residues for the APH but only 9 residues for the DCR, can be explained by incorporating the monomer's secondary structure. As the APH was shown to be helical[166, 167] it should form a helix with length of: $1.5\text{Å} \cdot N_{\text{residues}} = 31.5\text{Å}$; the same length the DCR could take, assuming a β-stranded structure: $3.5\text{Å} \cdot N_{\text{residues}} = 31.5\text{Å}$. Hence, both parts of the zipper motif could attach to each other and form a hairpin, spanning the hydrophobic inner part of a lipid bilayer.
The generated helical structure for the APH features two separated rows of charged residues instead of the suspected single row. The two spatially separated charge rows stop the DCR from simply lining up with the APH. Instead it is suggested that both charge rows are connected with two distinct DCRs forming intermolecular contacts between two neighboring TatA monomers. Guided by our MD simulations, we could determine the optimal zipper pattern for the self-assembly, which includes two inter- and five intramolecular contacts (Figure 4.6).

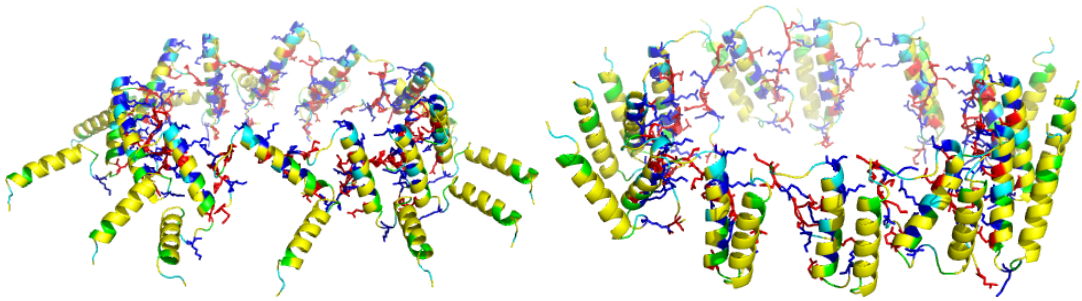
To model the self-assembly process of TatA_d pores, we set up isolated TatA_d monomers, equally distributed, around an implicit membrane pore (Section 4.2.2). Each monomer is positioned spatially separated from its neighbor molecules, but still connected pairwise by intermolecular salt bridges. Due to the implicit environmental approach, we could describe all systems within large simulation boxes (typically $200\text{Å} \times 200\text{Å} \times 200\text{Å}$) with no influence on the computational efficiency. All simulations comprise $200,000$ steps with a time step of $0.5fs$, resulting in a simulation time of $100ps$ in reduced units. Comparing with realistic times, a single TatA_d monomer would scarcely fold into its native conformation, but the reduced time of $100ps$ is sufficiently large to describe the whole pore-assembly process.
We commenced our investigation of the TatA_d pore complex with the easier description of the self-assembly process of tetrameric pores. Starting from four TatA_d subunits in either the open L-conformation (Figure 4.7b) or the more compact U-conformation (Figure 4.7b), we carried out 400 simulations with a pore diameter of 10Å and 100 simulations with a pore diameter of 9Å, respectively. The primarily chosen membrane thickness of 35Å seemed to be too large due

to the contribution of the additional 3Å transition regions on both membrane surfaces. Further $1,200$ simulations of the more promising U-conformation with a membrane thickness of 25Å and a larger energy contribution of the membrane-pore potential with a scaling factor of 5 resulted in well formed, compact pores.
Based on the optimal parameter set for tetrameric systems, we performed $1,000$ self-assembly simulations of dodecameric pore system around a 45Å cylindric membrane pore. However, we refined the parameters in $2,000$ simulations varying the pore radius (38Å, 40Å and 45Å), the transition region of the membrane-pore potential (3Å and 6Å), the scaling factor (5 and 8) and combinations thereof, to obtain more compact pores. We could achieve the best results with the same parameters for the transition regions (3Å) and the scaling factor (5). We chose 40Å as the pore diameter for further simulations.
Within $8,000$ simulations, we could determine the optimal salt-bridge pattern for the self-assembly of the dodecameric pores (Figure 4.6). The initially chosen pattern comprising a patch of five intramolecular salt bridges (E58-K24, K55-E27, E54-R30, E53-R34 and K52-E38) and a patch of two intermolecular contacts (residue$_i$-residue$_{i+1}$: K40-E51 and K44-D50) could perfectly fulfill all salt-bridge constraints (E27-K55: 2.2 ± 0.05Å, R30-E54: 2.2 ± 0.04Å, R34-E53: 2.2 ± 0.05Å, E38-K52: 2.3 ± 0.04Å, K40-E51: 2.3 ± 0.03Å, K44-D50: 2.3 ± 0.03Å), however the resulting complex seemed to fail to form a closed surface. A simple inversion of the contact affiliation to form five intermolecular (residue$_i$-residue$_{i+1}$: E58-K24, K55-E27, E54-R30, E53-R34 and K52-E38) and two intramolecular (K40-E51 and K44-D50) salt bridges did not produce adequate pore systems. However, a further change to the charge zipper motif, switching the sequential orientation of the intermolecular contacts of the patch of five intermolecular salt bridges to (residue$_i$-residue$_{i+1}$: K24-E58, E27-K55, R30-E54, R34-E53 and E38-K52), resulted in pore formations. Although the constraints could not be perfectly fulfilled, the observed pores featured a convincing closed surface.
Based on the final salt-bridge pattern, we could repeatedly simulate the formation of proper pore complexes, shaping an inner hydrophilic layer and a hydrophobic outer surface facing the membrane (Figure 4.7). The shape of the resulting pore complexes resembles a torus with an inner diameter of ≈ 40Å and a total width of ≈ 80Å. Both values are in very good agreement with experimental data gained from low resolution electron microscopy measurements of E.coli TatA complexes comprising twelve to 14 subunits[151]. The pore diameter matches the geometry of medium sized Tat substrate, which typically occurs in sizes ranging from 20Å to 70Å (Figure 4.7a)[168]. Further, a dynamic adaption of the number of participating monomers would be possible to perfectly fit the cargo. For the first time this would explain the observed diversity in pore sizes, as variably sized pores could assemble via zipping up a specific number of monomers[151, 169].

To support our proposed charge zipper motif, containing two intra- and five intermolecular contacts, we studied the influences of particular charge mutations on the TatA's capability to dimerize into active pores. This can be examined by measuring the ratio between monomers and oligomer in BN-PAGE (blue-native gel electrophoresis analysis)[170]. Ranging from $100kDa$ to $500kDa$, TatA appears in many homo-oligomeric complexes, which are proposed to be related to different sized pore assemblies[151, 171].



(a) *TatA pore: Starting and final conformation.*



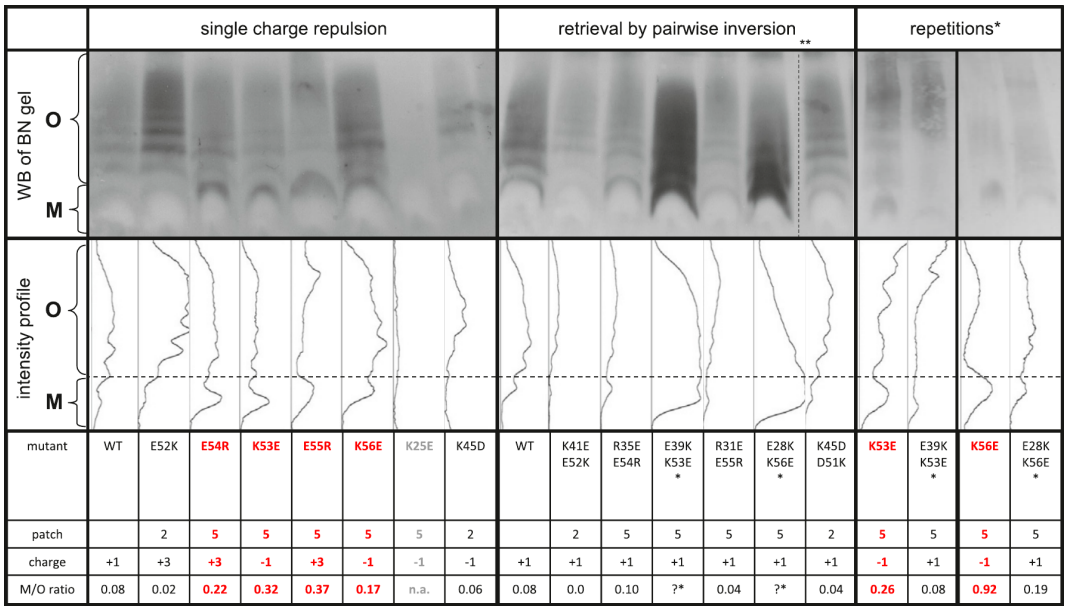
(b) *Assembled pores: L-form (left) and U-form (right).*

Figure 4.7.: *Starting from separated monomers (4.7a: left) we could repeatedly simulate the assembly of a dodecameric transmembrane pore. A qualitative overlay of a resulting pore with cryo-EM data (Copyright © (2008), American Chemical Society[151]) illustrates the matching proportions of our results (4.7a: right). We found two stable conformations for the twin arginine pore complex, a L-form (4.7b: left) and a U-form (4.7b: right). copyright © (2013), Elsevier[4]*

Our hypothesis of charge zippers being responsible for the formation of membrane complexes implies that an inversion of a single charge within the charge motif of TatA would prevent the monomers from forming stable, high-oligomeric complexes. Another charge mutation, aimed at the opponent residue to compensate the prior mutation, would restore the original equilibrium between monomers and oligomers. Further, charge mutations affecting intramolecular salt bridges should have less influences on this equilibrium, which allows us to distinguish them from intermolecular contacts.

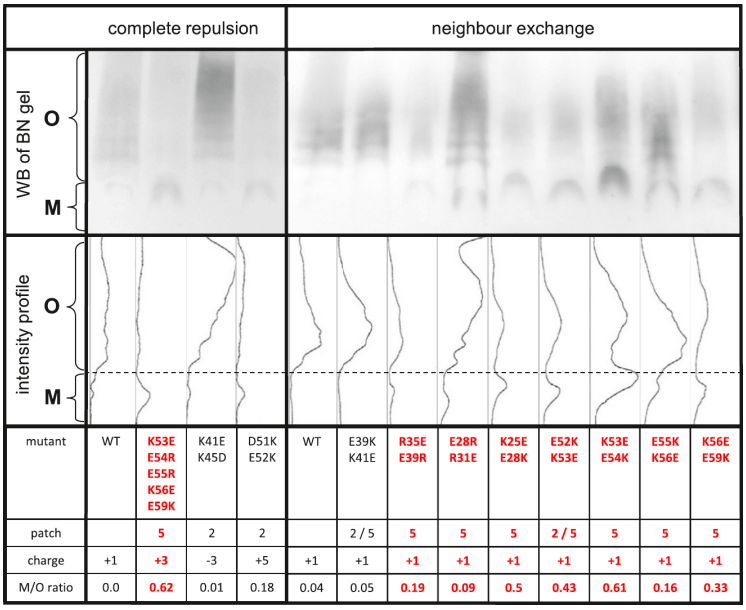
Extensive mutation studies carried out by our experimental partners comprise BN-PAGE analysis for both, single charge repulsion mutations, retrieval mutants and repetitions and complete repulsion mutations and neighbor exchange mutants (Figure 4.8). The ratio between the intensities of the monomeric and oligomeric bands compared to its value in the native wild type gives us a quantity for the degree of TatA disassembly. The summarized analysis for the mutation study is shown in figure 4.9[4].
Repulsion mutations within the patch were suspected to be engaged in intermolecular interactions between APH and DCR (K53E, E54R, E55R and K56E), resulted in increased disassembly of TatA. Repulsive mutations of K45D and E52K showed no influence on the oligomeric stability, which strengthens the hypothesis of five inter- and two intramolecular salt bridges. As the charge zippers would let us expect, the retrieval mutations, which additionally invert the opposed charge of each proposed salt bridge, could successfully restore the original stability of the oligomeric TatA states.
By collectively inverting all charge patterns within one patch at once, for either (K53E, E54R, E55R, K56E, E59K) or (K41E, K45D), the most clear effect on the oligomeric stability should arise. As expected, the complete repulsion mutations of the five-fold patch resulted in an essential disruption into monomers whereas the two-fold repulsion did not. However, they could also observe a different behavior of the two-fold patch by mutating the residue pair (D51K, E52K) on the DCR instead of (K41E, K45D) on the APH. Against common expectations, these mutations also destabilized the oligomeric state. Due to the high charge of the resulting TatA molecules, the oligomers could have become electrostatically unstable. So far, mutations changed the net charge of the wild-type from $+1$ to -1 or $+3$; an increase to $+5$ seems to disrupt the oligomerization process. To eliminate the influences of the changed net charge, further charge interchange mutations were carried out by exchanging neighboring charges: (E39K, K41E), (R35E, E39R), (E28R, R31E), (K25E, E28K), (E52K, K53E), (K53E, E54K), (E55K, K56E), and (K56E, E59K). Using the charge zipper force field model, seven of these mutations shifted the oligomeric equilibrium towards the monomeric state.
The experience gained from these mutational studies strengthens our hypothesis of structurally important charge zipper motifs being indispensable for the folding and self-assembly of flexible sized TatA translocation pores. However, the strict charge pattern is more likely to be a fuzzy zipper in which three salt bridges (K25:E59, E28:K56 and R35:E54) mainly make the intermolecular contacts and four residues (K41, K45, D51 and E52) act intramolecularly, potentially involving residues (K53 and E55) (Figure 4.6).

From our simulations, we suggest the DCR occurring in a flexible β-strand like structure that can easily attach to the APHs from either the left or the right side of the APH due to a flexible



mutant	WT	E52K	E54R	K53E	E55R	K56E	K25E	K45D	WT	K41E E52K	R35E E54R	E39K K53E *	R31E E55R	E28K K56E *	K45D D51K	K53E	E39K K53E *	K56E	E28K K56E *
patch		2	5	5	5	5	5	2		2	5	5	5	5	2	5	5	5	5
charge	+1	+3	+3	-1	+3	-1	-1	-1	+1	+1	+1	+1	+1	+1	+1	-1	+1	-1	+1
M/O ratio	0.08	0.02	0.22	0.32	0.37	0.17	n.a.	0.06	0.08	0.0	0.10	?*	0.04	?*	0.04	0.26	0.08	0.92	0.19

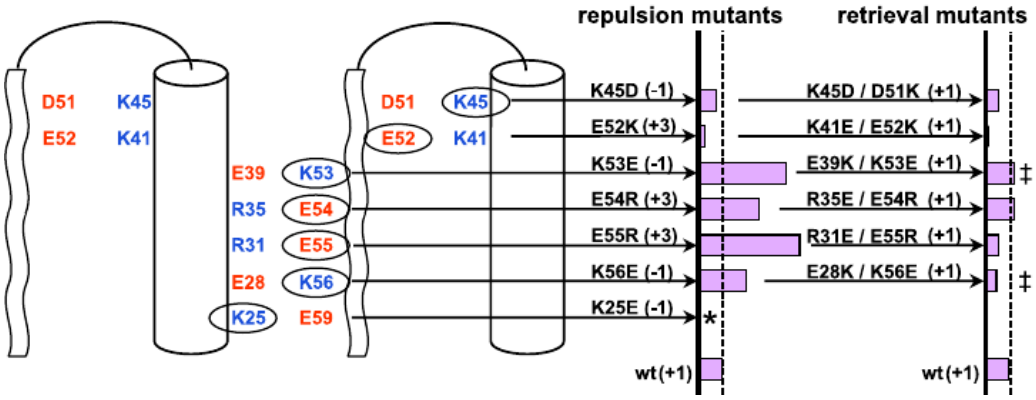
(a) *TatA: BN-PAGE part 1.*



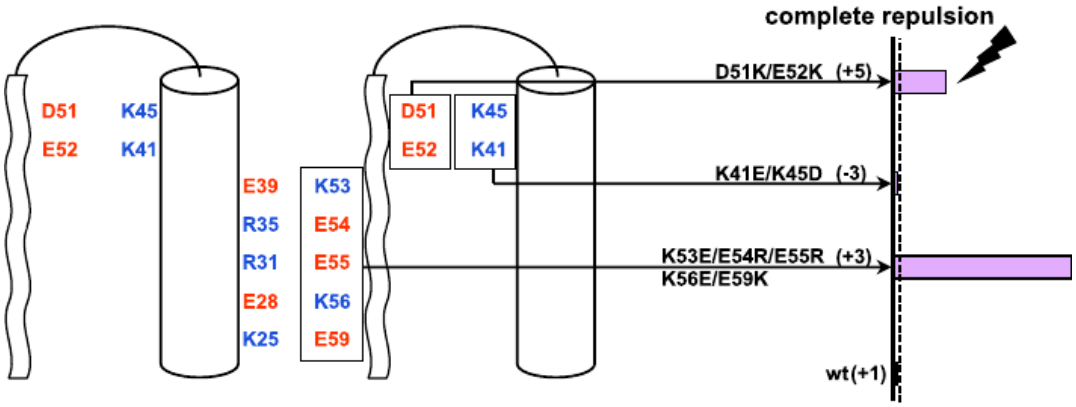
mutant	WT	K53E E54R E55R K56E E59K	K41E K45D	D51K E52K	WT	E39K K41E	R35E E39R	E28R R31E	K25E E28K	E52K K53E	K53E E54K	E55K K56E	K56E E59K
patch		5	2	2		2 / 5	5	5	5	2 / 5	5	5	5
charge	+1	+3	-3	+5	+1	+1	+1	+1	+1	+1	+1	+1	+1
M/O ratio	0.0	0.62	0.01	0.18	0.04	0.05	0.19	0.09	0.5	0.43	0.61	0.16	0.33

(b) *TatA: BN-PAGE part 2.*

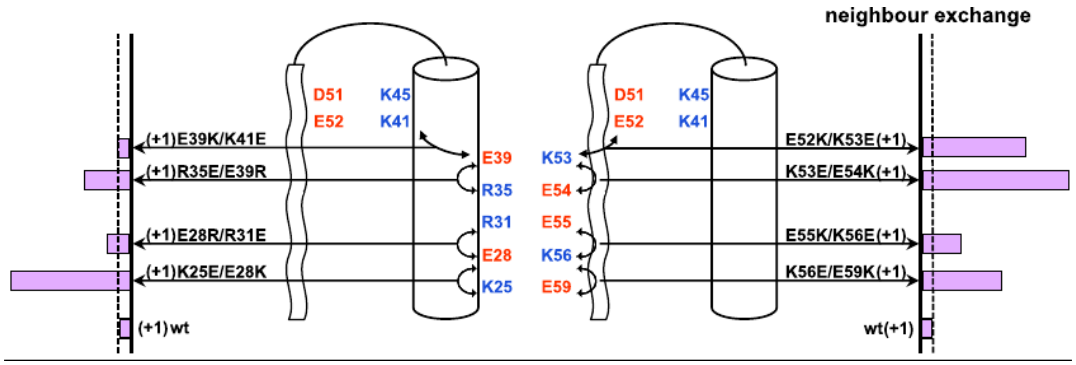
Figure 4.8.: *Using the BN-PAGE analysis, changes of the monomer/oligomer ratio induced by each mutation could be determined. A high M/O ratio indicates a resulting disassembly of the TatA complex. Mutations concerning the charge zipper motif, which destroyed the oligomeric stability, are highlighted in red. copyright © (2013), Elsevier[4].*



(a) *Single charge mutations.*



(b) *Complete repulsion mutations.*



(c) *Exchange mutations.*

Figure 4.9.: *Our TatA mutation study comprises single charge repulsion as well as retrieval mutations (4.9a), complete repulsion mutations (4.9b) and total charge maintaining exchange mutations (4.9c). copyright © (2013), Elsevier[4].*

$C_\beta - C_\gamma$ bond. However, in both ways multiple TatA monomers would form a concave wall with an adjustable curvature as the side chains involved in the salt-bridge formation are highly flexible. The resulting convex surface of the palisade then forms a hydrophobic shell, whereas the concave side carries the polar salt bridges. Assuming this flexible, amphiphilic palisade, the wall might arrange in two different ways in the membrane: resting flat on the membrane surface (inactive) or forming a curved transmembrane pore (active).

In agreement with a trap-door mechanism[168, 172–174] we propose that the active translocation pore, could be arranged by building these palisades. Single TatA units preassemble themself side by side, by concurrently being fixed in the lipid bilayer via their TMS helix[170, 175]. The resulting palisade would lie flat on the membrane layer, to either wait for the cargo protein to attach to or to combine itself with yet another palisade via their hairpins (front-to-front) or via their TMS helices (back-to-back).

Wrapping around the cargo or by forming a front-to-front trap-door, the flexible hinge region connecting TMS and APH would allow the hairpins to flip into the membrane. Two palisades could also cooperatively flip backward into the membrane keeping their L-form rigid (Figure 4.6). Both mechanisms would allow a dynamic arrangement of different sized pores fitted to the size of the associated cargo protein.

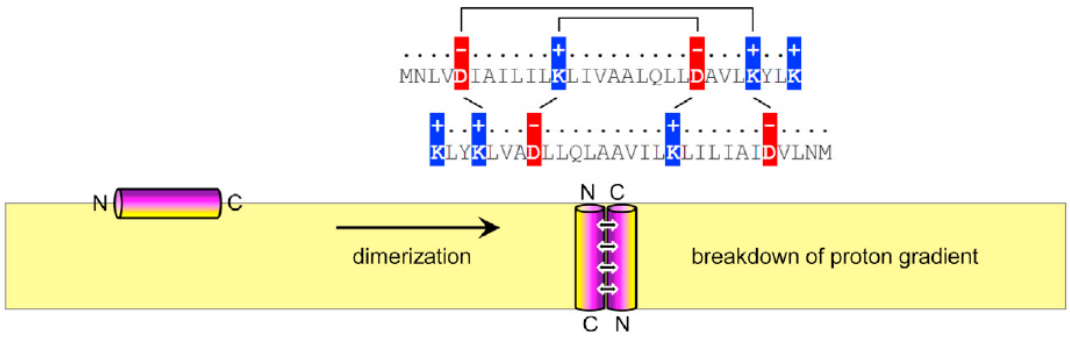
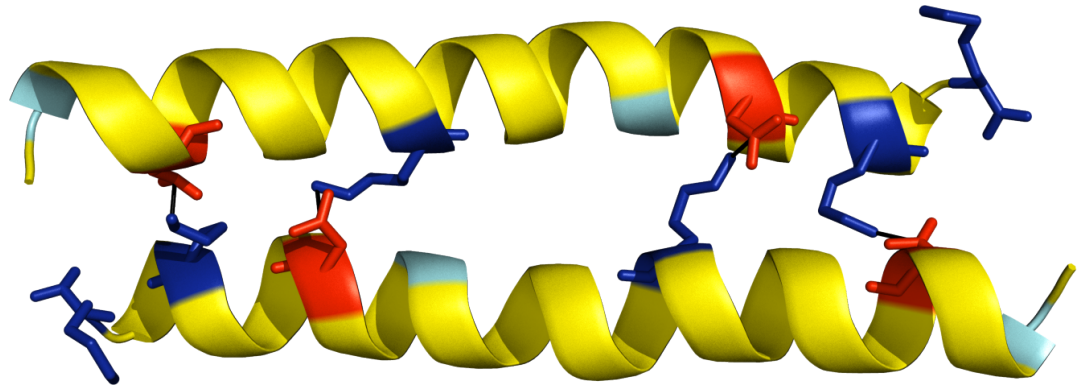


Figure 4.10.: *A dimeric structure of the bacterial stress-response peptide TisB could be stabilized through a ladder of four salt bridges allowing the dimer to span the membrane layer. copyright © (2013), Elsevier[4]*

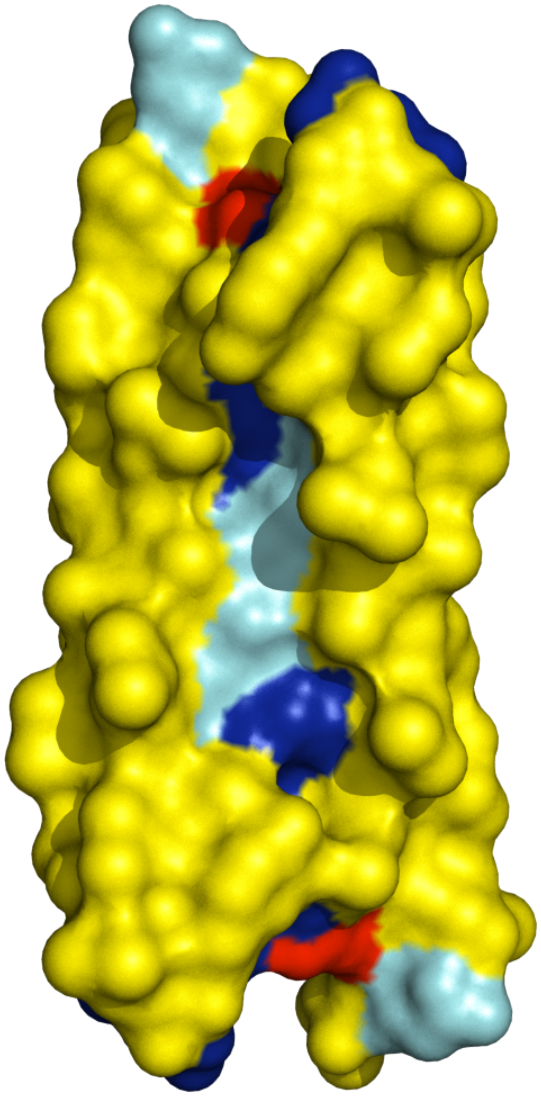
4.3.2. Bacterial Stress-Response Peptide: TisB

To survive the exposure to antibiotics, bacterial populations contain persister cells [176]. Persisters are mutations of the wild-type and make up only a tiny part of the whole population. Whereas, for growing cultures, in the exponential phase only few persisters are formed, in the stationary non-growing phase, they make up to $\approx 1\%$ of the population[177]. By residing in a dormant state, they resist currently available antibiotics thus ensuring the survival of the population[178]. Among other reasons, the generation of persisters can occur due to the involvement of toxin-antitoxin (TA) gene systems[179]. These gene systems often occur in multiple copies encoded in bacteria and archea and contain a toxin that downregulates cellular functionality and an antitoxin, which inactivates the toxin[180].
The TisB peptide is the toxic part of the TA system istR/tisAB[181] whose genes are induced as a SOS response to DNA-damaging antibiotics. It was shown that knocking out SOS-TA locus results in a decreased level of persister cells[182]. The TisB peptide shuts down the cell's metabolism by breaking down the proton motive force and decreasing ATP levels while being localized at the inner membrane[183].
Its structure could be determined by circular dichroism (CD) to be forming an amphiphilic helix whose relative orientation to the membrane layer was measured using oriented circular dichroism (OCD). The TisB monomer contains an intrinsic charge pattern that suggests antiparallel dimerization via four intermolecular salt bridges (Figure 4.10).

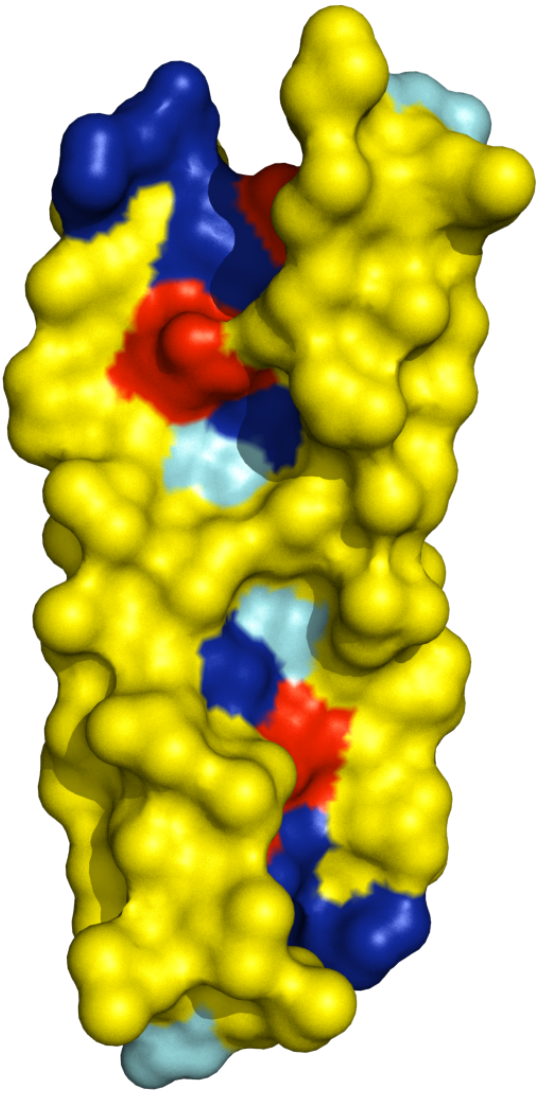
Assuming four intermolecular salt bridges between two TisB monomers, we could attach both helices close to each other without breaking any secondary structure (Figure 4.11). Within 100 simulations of $100ps$ each could describe the self-assembly of TisB dimers inside a 35Å membrane. In the assembled configuration all four constraints are fulfilled (D5-K26: 2.4Å, K12-D22: 2.4Å, D22-K12: 1.9Å and K26-D5: 1.8Å). The surface representation reveals a mostly apolar outer surface of the dimer with all hydrophobic residues buried between both helices (Figure 4.11b and 4.11c). To bury the hydrophobic regions within the inner of the structure, both helices have to twist around each other. The mostly apolar transmembrane segment could be easily inserted into the membrane and it is stabilized in this position by both polar termini. The ladder of charged residues resides inside the dimer and could be responsible for breaking down the membrane's protein gradient.



(a) *Constraints.*



(b) *Surface: Front.*



(c) *Surface: Back.*

Figure 4.11.: *A dimeric structure of two TisB helices can be assembled by forming a patch of four salt bridges. The structure of the resulting dimer is sterically feasible and would potentially permit the insertion into a membrane.*

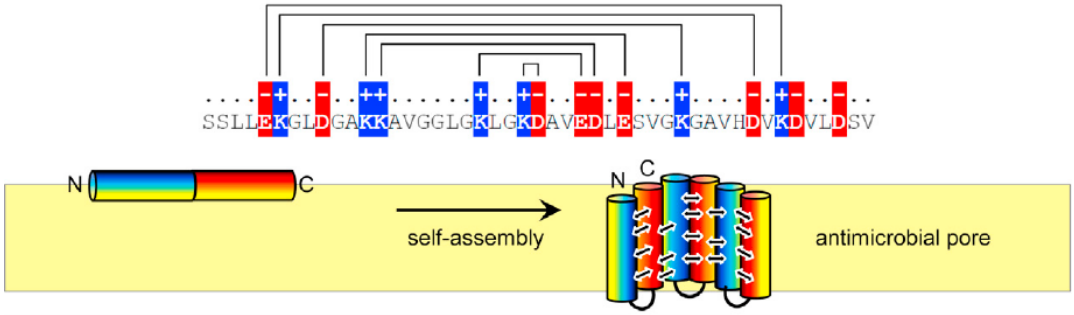


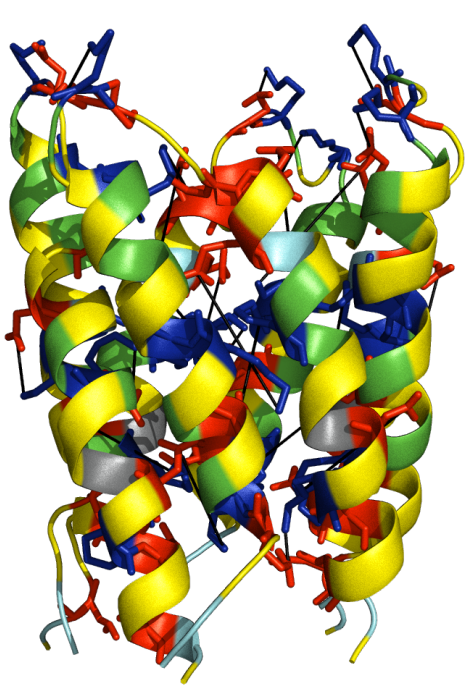
Figure 4.12.: *The sequence of the anionic antimicrobial peptide Dermcidin reveals an intrinsic alternating charge pattern suggesting a hairpin like self-assembly. By forming intermonomeric salt bridges a stable pore complex can be assembled. copyright © (2013), Elsevier[4]*

4.3.3. Anionic Antimicrobial Peptide: Dermcidin

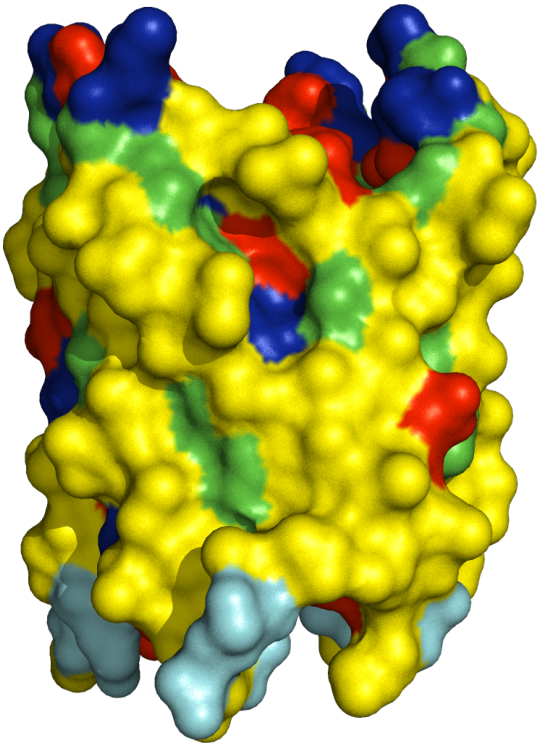
For many species, antimicrobial peptides (AMP) are an important element of the innate immune response[184]. AMPs reside in the epithelia of many organisms, including mammals. Within the first few hours after an injury of the epithelia and whilst wound healing, they control microbial growth [185]. One distinguishes between cathelicidins[186], ingredients of the wound fluid, and defensins[187], small cationic peptides (CAMP), active against gram- positive and negative bacteria. However, due to co-evolution of CAMPs and CAMP-resistance mechanisms, bacteria have evolved resistance mechanism against CAMPs[188]. The Dermcidin gene was found in sweat glands and encodes anionic antimicrobial peptides (AAMP) with no homology to previous genes[185]. Dermcidin peptides maintain their activity inside the epidermal surface even at high salt concentrations and within a wide pH spectrum. It is thought that these very rare AAMPs have evoled as a response to CAMP-resistance of bacteria[189].
Oriented CD measurements revealed that DCD-1L, the most widely studied Dermcidin, binds to the membrane surface but can also form pores, which is a very common property of antimicrobial peptides. This enables the formation of ion channels through the bacterial membrane induced by the Zn^{2+} concentration[189]. The DCD-1L helix length equals twice the thickness of the lipid layer. By forming a helical hairpin stabilized through intra- and intermolecular salt bridges, an oligomeric DCD-1L structure can self-assemble and form an antimicrobial pore (Figure 4.12).

We could demonstrate that single Dermcidin monomers are in principle able to form a hairpin like structure by forming a ladder of salt bridges, which would then allow them to assemble into oligomeric pore complexes. By proposing a salt-bridge pattern comprising five intra- (E5-K41, D9-K34, K12-E30, K20-E27, K23-D24) and two intermolecular contacts (K6-D39, K13-D28) within a complex of four DCD-1L monomers, we were able to simulate the assembly process of a tetrameric transmembrane pore standing upright in the membrane. The assembled complex comprises a highly charged inner core, which is shielded by an apolar outer surface facing the inner of the membrane (Figure 4.13c). The apolar surface of the pore, which is directed to the inner of the membrane would permit a stable positioning of the complex. Within the inner core, a small charged channel emerges, which could facilitate the experimentally observed ion transport across the membrane.

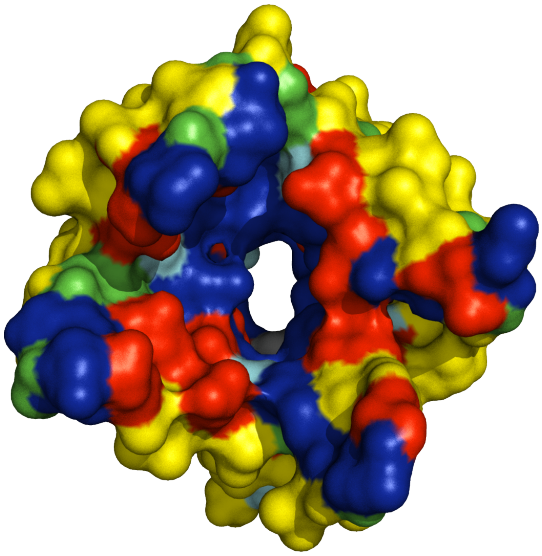
In contrast to our hairpin-shaped subunits, a recent study proposed a hexameric helix bundle structure consisting of six monomers, each one shaped as a continuous long helix[17]. In this scenario, the helix bundle lies tilted in the membrane with an angle of $\approx 30^{\circ}$ and facilitates ion transport through lateral holes in the hexamer.



(a) *Dermcidin pore: Constraints.*



(b) *Dermcidin pore: Front.*



(c) *Dermcidin pore: Top.*

Figure 4.13.: *The assembled pore complex consisting of four distinct Dermcidin monomers features the salt-bridge pattern as well as the observed helical structure (left). The outer pore surface is mostly hydrophobic (middle), whereas the inner channel surface contains charged amino acids.*

4.3.4. Structural Glycoprotein of Pestiviruses: E^{rns}

The group of pestiviruses comprises four species: both bovine viral diarrhea viruses (BVDV-1/2), the classical swine fever virus (CSFV) and the border disease virus of sheep (BDV)[190]. They belong to the positive strand RNA-viruses[191] and have high economic impact, by being contracted by farm animals. Based on their molecular structure, they share some similarities with the human hepatitis C virus[192] with the obvious difference that its genome includes coding sites for two further proteins: An unstructured protein N^{pro} and the E^{rns} envelope protein. The E^{rns} peptide, a part of the RNase T_2 superfamily[193, 194], is involved in ribonuclease activity[195, 196] and can trigger a neutralizing antibody response of the host[197]. It lacks the typical transmembrane region, where the retention signal is located for all flaviviruses[190]. Here, the formation of a C-terminal amphipathic helix, which binds via a membrane anchor to the membrane surface of the infected host cell, allows to serve as a retention signal[191, 198].
To interfere with intracellular RNA using its RNase capabilities, E^{rns} would have to reside within the cytosol or even at the nucleus, which would require a translocation through the cell membrane of the target cell[198]. The initial binding of E^{rns} to the cell may happen through interactions with glycosaminoglycans (GAGs)[199–201] and it was shown that dimeric E^{rns} structures could translocate into cells and further could be able to transport large enzymes into the cell[202].
As other virus-encoded glycoproteins, E^{rns} could play an important role for the development of effective vaccines[203]. Removal of the E^{rns} component from the viral genome does not impair its capability for autonomous RNA replication, however it could no longer create infectious virions[204, 205]. Mutated or knocked out E^{rns}-coding sequences were shown to be clinically attenuated[206–208].

The sequence of the E^{rns} peptide comprises a pronounced ladder of seven charge zipper pairs (E4-K66, D9-K64, K13-E62, E14-K60, D16-K59, D23-R52, D30-K45 and E37-R40 (Figure 4.14)). We propose that a self-assembled E^{rns} structure could bury hydrophilic residues between two formed helices to penetrate the membrane. In simulations without zipper constraints, the E^{rns} peptide aligns to the membrane surface by forming two faces: a hydrophilic and a hydrophobic. This behavior allows a perfect alignment of the hydrophobic faces to the membrane surface. For the simultaneous alignment of both helical parts, a kink of 140° develops between them. After switching on the salt-bridge potential, E^{rns} instantaneously forms a hairpin like structure, which also aligns to the membrane surface with by forming a hydrophobic surface (Figure 4.15). Except for the two contacts K13-E62 and E14-K60, all other constraints could be formed. However, only a small rotation of the kink angle would be necessary to bury the majority of hydrophilic amino acids within the hairpin, which would then permit the membrane insertion. A further option for shielding polar residues could be to form a dimer by attaching a second E^{rns} hairpin.

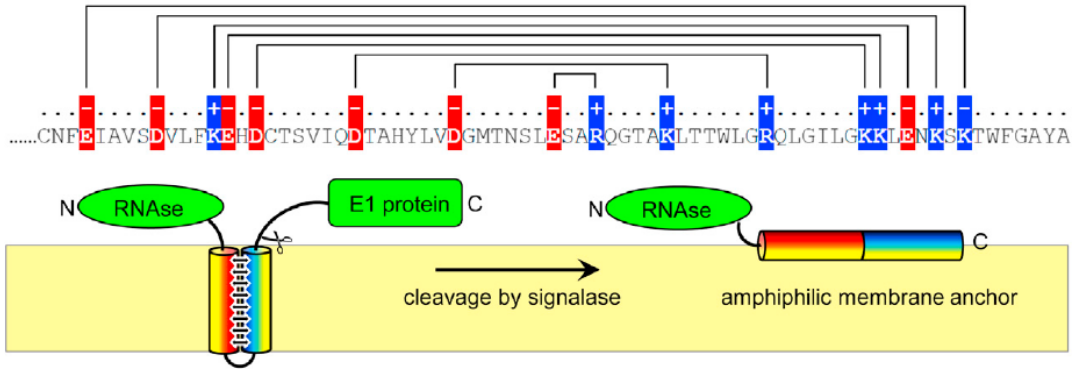
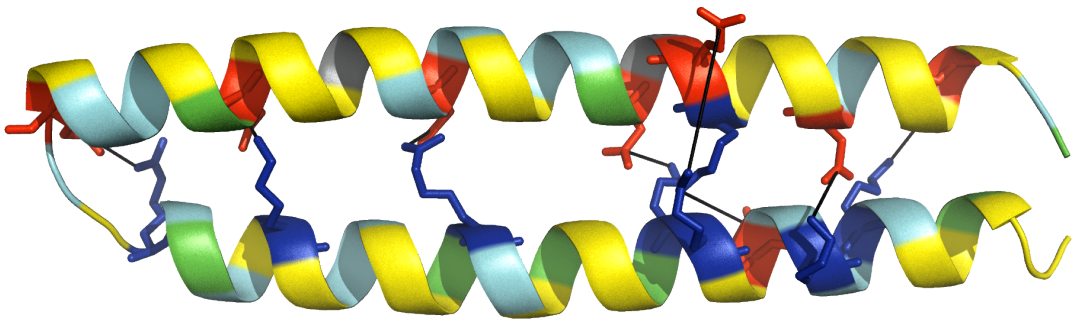
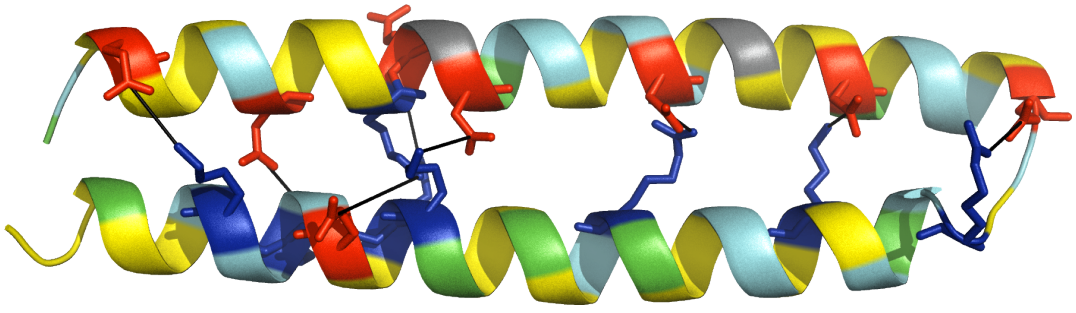


Figure 4.14.: *The sequence of the structural glycoprotein of pestiviruses E^{rns} also includes charge zipper motif which suggests a hairpin like assembly for the formation of transmembrane structures. copyright © (2013), Elsevier[4]*



(a) E^{rns} *: Front.*



(b) E^{rns} *: Back.*

Figure 4.15.: *The hairpin like structure allows the E^{rns} protein to assemble a ladder of eight salt bridges. The resulting structure exhibits a hydrophilic surface and an apolar face to attach to the membrane layer. Either a rotation of the kink angle could bury charges between both helices or a second attached monomer could shield charges, to permit membrane insertion of the otherwise amphipathic helix.*

4.4. Discussion

In collaboration with the workgroup Ulrich at the KIT and the group Ruggerone at the Università di Cagliari we proposed a novel mechanism for the driving force behind protein self-assembly processes in the proximity or inside membranes[4]. We have identified intrinsic complementary charge patterns within the primary sequence of membrane proteins and investigated their role in the development of the three dimensional structure of the associated protein complexes. For the $TatA_d$ protein of the twin arginine translocation complex, this charge pattern manifests itself as a zipper of charges between adjacent structural parts. We suggested that the stability of a single $TatA_d$ monomer is achieved by successively forming salt bridges in a zipper like manner as well as intermolecular salt bridges to stabilize the oligomeric structure. To demonstrate the steric feasibility of the salt bridges, we developed an implicit membrane-pore force field within a structure based potential, which we could use to successfully describe the assembly process of dodecameric $TatA_d$ pores. In an extensive mutational study, our experimental partners could verify the salt-bridge pattern and assign it to four intra- and three intermolecular contacts. The geometry of the simulated pores matches well with existing cryo-EM measurements of assembled pore systems. Since TatA forms heterogeneous complexes, it could not yet be resolved with high resolution and atomistic data only exists for single monomers or fragments thereof.

5. Multiple Try Metropolis

5.1. Motivation

Both thermodynamic sampling methods described so far, molecular dynamics and the Metropolis Hastings method, are strictly sequential algorithms, meaning that they are based on successive energy evaluations. Without changing the complexity of the force field, the sampling speed can only be improved by code engineering. A common approach to parallelize the energy evaluation of an MD simulations is domain decomposition[209]. By spatial decomposing the simulation volume, a simulation can be distributed onto multiple CPUs on shared memory architectures. However, this approach is usually limited to only a few numbers of CPUs[107] and, like other parallelization strategies to non-bonded interactions (Section 3.4.1). Domain decomposition is not readily available for implicit solvent simulations with a small amount of atoms. If a decomposition cell contains less than 100 atoms, there is a large overhead due to load imbalance caused by the "statistical fluctuation of the number of particles in a domain decomposition cell"[107]. Instead of enhancing a single simulation, multilevel sampling and optimization methods have been established, which carry out several distinct simulations at once, allowing interchanges after certain (time-)steps [116, 117]. Although these methods enhance the overall phase space traversal, they cannot accelerate a single simulation.

The reliance on basic physics principles is a major benefit of MD simulations[57] and permits the simulation of deterministic, time-resolved trajectories. However, the integration of Newton's equations of motion is a sequential process and cannot be parallelized. The time scale problem is caused by the discrepancy between the femtosecond integration step of MD and the investigation of long scale processes, occurring on time scales beyond the microsecond scale.
But also the stochastic MH method is a serial process and thus can only be accelerated by parallelizing the energy evaluation. MH simulations are not restricted to a maximal transition step size $\|T(\vec{x}_{old} \rightarrow \vec{x}_{new})\|$; however choosing the transition step size too wide results in low acceptance ratios α_{MH} and therefore leads to ineffective algorithms[57, 210]. On the other hand, an insufficiently small transition step size leads to a local search and thus results in a slow converging algorithm, which can easily be trapped in local minima. Several approximations have been developed to overcome this problem; yet with higher dimensionality of the system their accuracy gets worse or they become too expensive[211].

The Multiple Try Metropolis method (MTM) has been proposed as a promising modification to the MH algorithm that proposes multiple trial configurations per step. Studies on simplified models observed significant improvements in comparison to common MH simulations[210]. In the following, we illustrate the MTM method, the benefits it grants and finally its applicability to high dimensional systems such as the potential function of a protein.

5.2. Methods

5.2.1. Multiple Try Metropolis

The Multiple Try Metropolis method extends the MH method in a way that in each proposal step a fixed number of trials is generated. Due to an modified acceptance criterion the detailed balance is attained and hence the MTM induces a reversible Markov chain with a stationary distribution $\pi(\vec{x})$[210]. For this purpose, the MTM requires a second proposal step, starting from the prior selected trial configuration. Nonetheless, an MTM approach is computationally more efficient than using a conventional MH method. Assuming the system is currently located in the state $\vec{x}_t$; a single MTM transition is defined as:

1. Generate N trials: $\vec{y}_1, ..., \vec{y}_k$ using the proposal transition function $T(\vec{x}_t \to \cdot)$ and compute $\omega(\vec{x}_t \to \vec{y}_i)$ for all $i = 1, ..., K$.

   For each trial, we define the corresponding transition probability:

   $$\omega(\vec{x} \to \vec{y}) = \pi(\vec{y})T(\vec{x} \to \vec{y})\lambda(\vec{x}, \vec{y}) \qquad , \tag{5.1}$$

   where $\lambda(\vec{x}, \vec{y})$ is a non-negative, symmetric function in $\vec{x}$ and $\vec{y}$ with the requirements that:

   $$\lambda(\vec{x}, \vec{y}) > 0 \quad \text{whenever} \quad T(\vec{x} \to \vec{y}) > 0$$

   and $T(\vec{x} \to \vec{y})$ is restricted to:

   $$T(\vec{x} \to \vec{y}) > 0 \quad \text{if and only if} \quad T(\vec{y} \to \vec{x}) > 0 \qquad .$$

2. Select one state $\vec{y}$ among the whole trial set $\{\vec{y}_i\}$, with $i = 1, ..., N$ proportional to their probabilities: $\omega(\vec{x} \to \vec{y}_i)$.

3. Draw $N - 1$ new trials $\vec{x}_1^*, ..., \vec{x}_{N-1}^*$ from the distribution $T(\vec{y} \to \cdot)$ starting from the just selected state $\vec{y}$ and let the N-th state be $\vec{x}_N^* = \vec{x}_t$.

4. Accept the proposed transition of the system into the state $\vec{y}$ with the probability:

   $$\alpha_{MTM} = \min\left\{1, \frac{\sum_{i=1}^{N} \omega(\vec{x}_t \to \vec{y}_i)}{\sum_{i=1}^{N} \omega(\vec{y} \to \vec{x}_i^*)}\right\}$$

5. Update the configuration $\vec{x}_{t+1} \equiv \vec{y}$ provided the step was accepted, otherwise reside in the current state $\vec{x}_{t+1} \equiv \vec{x}_t$ once more.

The stepsize $T(\vec{x} \to \vec{y})$, the number of trials N and the shape of the distribution $\pi(\vec{x})$ have strong influences on the efficiency of the MTM method[56]. Also the modification of the symmetric function $\lambda(\vec{x}, \vec{y})$ changes the behavior of the MTM; the two most interesting cases are:

- MTM-inv: $\lambda(\vec{x}, \vec{y}) = \{T(\vec{x} \to \vec{y})T(\vec{y} \to \vec{x})\}^{-1}$

- MTM-I: $\lambda(\vec{x}, \vec{y}) = 1$

5.2.2. Generalized Multiple Try Metropolis

The Generalized Multiple Try Metropolis (GMTM) uses less restrictive selection probabilities than the MTM (Section 5.2.1) without violating the detailed balance condition[56]. Its schematic functionality is visualized in Figure 3.9 and follows the stepwise application of a MTM-like simulation scheme:

1. Generate N trials $\vec{y}_1, ..., \vec{y}_N$ using the proposal transition function $T(\vec{x}_t \to \cdot)$ and compute $\omega^*(\vec{x}_t \to \vec{y}_i)$ for all $i = 1, ..., N$, where $\omega^*(\vec{x} \to \vec{y})$ is an arbitrary function satisfying:

$$\omega^*(\vec{x} \to \vec{y}) > 0 \qquad .$$

2. Select one state $\vec{y}$ among the whole trial set $\{\vec{y}_i\}$, with $i = 1, ..., N$ proportional to their probability:

$$p_{\vec{y}_i} = \frac{\omega^*(\vec{x}_t \to \vec{y}_i)}{\sum_{j=1}^{N} \omega^*(\vec{x}_t \to \vec{y}_j)} \qquad .$$

3. Draw $N-1$ new trials $\vec{x}_1^*, ..., \vec{x}_{N-1}^*$ from the distribution $T(\vec{y} \to \cdot)$ starting from the just selected state $\vec{y}$ and let the N-th state be $\vec{x}_N = \vec{x}_t$.

4. Define the inverse probability:

$$p_{\vec{x}_t} = \frac{\omega^*(\vec{y} \to \vec{x}_t)}{\sum_{j=1}^{N} \omega^*(\vec{y} \to \vec{x}_j^*)} \qquad .$$

5. Accept the proposed transition of the system into state $\vec{y}$ with the probability:

$$\alpha_{GMTM} = \min\left\{1, \frac{\pi(\vec{y})T(\vec{y} \to \vec{x}_t)p_{\vec{x}_t}}{\pi(\vec{x}_t)T(\vec{x}_t \to \vec{y})p_{\vec{y}}}\right\} \qquad .$$

6. Update the configuration $\vec{x}_{t+1} \equiv \vec{y}$ provided the step was accepted, otherwise reside in the current state $\vec{x}_{t+1} \equiv \vec{x}_t$ once more.

Analog to the MTM method, we can again define the two special GMTM cases:

- GMTM-inv: $\omega^*(\vec{x}_t \to \vec{y}_i) = \pi(\vec{y}_i)T(\vec{x}_t \to \vec{y}_i)$
- GMTM-I: $\omega^*(\vec{x}_t \to \vec{y}_i) = \frac{\pi(\vec{y}_i)}{T(\vec{y}_i \to \vec{x}_t)}$

5.2.3. Model System

Contrary to common intuitions build on MH based simulations, MTM transition probabilities can behave completely different, because they strongly depend on the local densities of states

surrounding each point in the phase space. Even in simple one dimensional energy landscapes, we could find scenarios where a standard MTM seems to behave non-physically. For the understanding of how a MTM method actually behaves, a one dimensional, discrete model system is a demonstrative example as every transition rate, energy level as well as the distribution of states can be calculated analytically.
Assuming a linear energy landscape with equidistantly distributed states $E_i = e^{-i\beta\Delta}$ where $i \in \{-n, 1-n, \ldots, 0, \ldots, n-1, n\}$ and $\Delta \in \{\mathbb{R} \setminus 0.0\}$. We remove the exponential functions by rewriting all energies E_i in potentials of $E_1 = \chi \equiv e^{-\beta\Delta}$. The transition ratio between two adjacent states can be determined by calculating the sum over all possible MTM pathways which jump from state E_0 to state E_1 without loss of generality because Δ can either by positive or negative.

At the first stage of MTM method we generate N new trial states $\vec{y}_i$ starting from the current state $\vec{x_t}$ using the transition function $T(\vec{x}_t \to \vec{y}) \equiv T : \vec{y} \to \vec{x}_t \pm \Delta$. Starting at state $\vec{x}_t \mathrel{\widehat{=}} E_0$, this leads to a trial set $\{\vec{y}\}$ including u times χ and $(N-u)$ times χ^{-1}. The probability to now pick one of the states χ is given by:

$$p_\chi = \frac{u\chi}{u\chi + (N-u)\chi^{-1}} \tag{5.2}$$

The MTM acceptance probability for state $\vec{y}$ is given by the energy distribution in the second stage. Starting from state $\vec{y} \mathrel{\widehat{=}} \chi$ we use the transition function $T(\vec{y} \to \vec{x}^*)$ to generate N new states $\{\vec{x}_0^*, \ldots, \vec{x}_N^*\}$ and let $\vec{x}_N^* = \vec{x}_t$. This new set contains v times χ^2 and $(N-v)$ times $\vec{x}_t \mathrel{\widehat{=}} 1$. The resulting acceptance probability for state $\vec{y}$ is defined by:

$$\alpha = \min\left\{1, \frac{u\chi + (N-u)\chi^{-1}}{(N-v) + v\chi^2}\right\}$$

To achieve the overall transition probability Π_{01} we need to scale every one of these pathways by its stochastic probability. The Binomial distribution $P_{\text{binomial}}(N, k, p)$ gives the probability of drawing an event, which has the probability p to show up, k times out of N total draws. As we only allow nearest neighbor jumps, $p = 0.5$ for all processes.

$$P_{\text{binomial}}(N, k, p) = \binom{N}{k} p^k \cdot (1-p)^{N-k}$$

We now can write the transition probability $\Pi_{1\chi}$ as:

$$\begin{aligned}
\Pi_{1\chi} &= \sum_{u=0}^{N} P_{\text{binomial}}(N,u,p)\cdot p_{\chi}\cdot\left[\sum_{v=0}^{N-1} P_{\text{binomial}}(N-1,v,p)\cdot\alpha\right] \\
&= \sum_{u=0}^{N}\binom{N}{u}p^{u}\cdot(1-p)^{N-u}\cdot\frac{u\chi}{u\chi+(N-u)\chi^{-1}}\cdot \\
&\quad\left[\sum_{v=0}^{N-1}\binom{N-1}{v}p^{v}\cdot(1-p)^{N-v-1}\cdot\min\left\{1,\frac{u\chi+(N-u)\chi^{-1}}{(N-v)+v\chi^{2}}\right\}\right] \\
&= \left(\frac{1}{2}\right)^{2N-1}\cdot\sum_{u=0}^{N}\binom{N}{u}\cdot\frac{u\chi}{u\chi+(N-u)\chi^{-1}}\cdot \\
&\quad\left[\sum_{v=0}^{N-1}\binom{N-1}{v}\cdot\min\left\{1,\frac{u\chi+(N-u)\chi^{-1}}{(N-v)+v\chi^{2}}\right\}\right] \\
&= \left(\frac{1}{2}\right)^{2N-1}\cdot\sum_{u=0}^{N}\sum_{v=0}^{N-1}\binom{N}{u}\binom{N-1}{v}\cdot\min\left\{1,\frac{u\chi}{(N-v)+v\chi^{2}}\right\} \quad .
\end{aligned} \tag{5.3}$$

As both transition probabilities are symmetric, the $\Pi_{\chi 1}$ can be obtained by inverting χ within $\Pi_{1\chi}$. With a trial size of $N = 1$, the MTM reduces to a Metropolis Hastings approach and equation 5.3 simplifies to:

$$\begin{aligned}
\Pi_{1\chi} &= \frac{1}{2}\cdot\sum_{u=0}^{1}\binom{1}{u}\cdot\frac{u\chi}{u\chi+(1-u)\chi^{-1}}\cdot \\
&\quad\left[\sum_{v=0}^{0}\binom{0}{v}\cdot\min\left\{1,\frac{u\chi+(1-u)\chi^{-1}}{(1-v)+v\chi^{2}}\right\}\right] \\
&= \frac{1}{2}\left[0+\min\left\{1,\chi\right\}\right] \\
&= \frac{1}{2}\min\left\{1,e^{-\beta\Delta}\right\} \quad ,
\end{aligned}$$

and the acceptance ratio converges against the Metropolis acceptance criterion.

The compliance with the detailed balance can now easily be shown:

$$\begin{aligned}
\Pi_{1\chi}\cdot p_1 &= \Pi_{\chi 1}\cdot p_{\chi} \\
\Pi_{1\chi} &= \Pi_{\chi 1}\cdot\chi \\
\min\left\{1,e^{-\beta\Delta}\right\} &= \min\left\{1,e^{\beta\Delta}\right\}\cdot e^{-\beta\Delta} \\
&= \min\left\{e^{-\beta\Delta},1\right\} \\
&\quad q.e.d. \quad .
\end{aligned}$$

As an example, we demonstrate the analytic solution of the transition probabilities for most trivial MTM approach with a trial size of $N = 2$ that includes transition list holding two

states:

$$\begin{aligned}
\Pi_{1\chi} &= \frac{1}{8}\cdot\sum_{u=1}^{2}\binom{2}{u}\cdot\frac{u\chi}{u\chi+(2-u)\chi^{-1}} \\
&\quad\cdot\sum_{v=0}^{1}\binom{1}{v}\cdot\min\left\{1,\frac{u\chi+(2-u)\chi^{-1}}{(2-v)+v\chi^{2}}\right\} \\
&= \frac{1}{8}\left[\frac{2\chi}{\chi+\chi^{-1}}\left(\min\left\{1,\underbrace{\frac{\chi+\chi^{-1}}{2}}_{=\cosh(\pm\beta\Delta)\geq 1}\right\}+\min\left\{1,\frac{\chi+\chi^{-1}}{1+\chi^{2}}\right\}\right)\right. \\
&\quad\left.+\min\{1,\chi\}+\min\left\{1,\underbrace{\frac{2\chi}{1+\chi^{2}}}_{=\cosh^{-1}(\pm\beta\Delta)<1}\right\}\right] \\
&= \frac{1}{8}\left[\chi\cosh^{-1}(\beta\Delta)\left(1+\min\left\{1,\chi^{-1}\right\}\right)+\min\{1,\chi\}+\cosh^{-1}(\beta\Delta)\right] \\
\Pi_{\chi 1} &= \frac{1}{8}\left[\chi^{-1}\cosh^{-1}(\beta\Delta)\left(1+\min\{1,\chi\}\right)+\min\left\{1,\chi^{-1}\right\}+\cosh^{-1}(\beta\Delta)\right] \qquad .
\end{aligned}$$

If we now insert the simplified transition probabilities into the detailed balance criterion we can again show the compliance. The more complex detailed balance criterion now includes all possible transition pathways between both states:

$$\begin{aligned}
\Pi_{1\chi}\cdot p_1 &= \Pi_{\chi 1}\cdot p_\chi \\
\Pi_{1\chi} &= \Pi_{\chi 1}\cdot\chi \\
\frac{1}{8}\left[\chi\cosh^{-1}(\beta\Delta)\left(1+\min\left\{1,\chi^{-1}\right\}\right)\right. &= \frac{\chi}{8}\left[\chi^{-1}\cosh^{-1}(\beta\Delta)\left(1+\min\{1,\chi\}\right)\right. \\
\left.+\min\{1,\chi\}+\cosh^{-1}(\beta\Delta)\right] &\quad \left.+\min\left\{1,\chi^{-1}\right\}+\cosh^{-1}(\beta\Delta)\right] \\
\chi+\chi\min\left\{1,\chi^{-1}\right\} &= 1+\min\{1,\chi\} \\
+\cosh(\beta\Delta)\min\{1,\chi\}+1 &\quad +\chi\cosh(\beta\Delta)\min\left\{1,\chi^{-1}\right\}+\chi \\
\min\left\{1,\chi^{-1}\right\}(\chi-\chi\cosh(\beta\Delta)) &= \min\{1,\chi\}(1-\cosh(\beta\Delta)) \\
\frac{\chi(1-\cosh(\beta\Delta))}{(1-\cosh(\beta\Delta))} &= \frac{\min\{1,\chi\}}{\min\{1,\chi^{-1}\}} \\
\chi &= \frac{\min\{1,\chi\}}{\min\{1,\chi^{-1}\}} \\
q.e.d. &\qquad .
\end{aligned}$$

The compliance with the detailed balance criterion implies that a stochastic method, applied to a potential function, will result in a distribution π which equals the systems stationary distribution with converging accuracy. A general proof that the MTM satisfies the detailed balance condition can be found here[57]. Having a working stochastic method allows us now to uncover the influences of the arbitrary parameters: λ, N, Δ and the complexity of the underlying potential surface. In our studies, we could uncover two major issues concerning the robustness of the MTM method. We could find certain scenarios, where the efficiency of the method dramatically

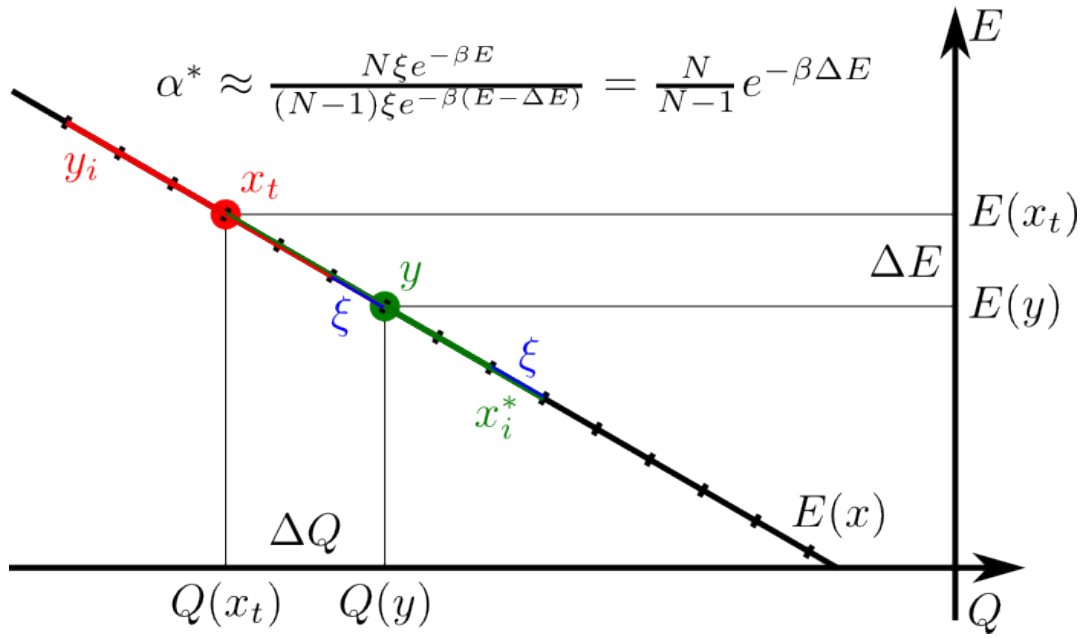


Figure 5.1.: *Assuming a linear potential surface and an equal distributed* $T(\vec{x}_t \to \vec{y}) \equiv T : \vec{y} \to \vec{x}_t + i\frac{\Delta Q}{2}$*, with* $i \in [-1, -1 + \frac{2}{N}, ..., 1 - \frac{2}{N}, 1] \backslash 0$*, the MTM will pick a* $y \mathrel{\widehat{=}} E_{x_t} - \Delta E$*. The energy distribution of the second trial ensemble, starting from* y*, will then contain the same energies, a factor* ΔE *lower. The acceptance ratio for the proposed state* y *will then be dominated by only a fraction* ξ *of states. The resulting* $\alpha = \min(1, \alpha^*)$ *decreases with increasing trials size* N *and roughness* ΔE*.*

decreases. The first one depends on the roughness of the potential surface and the second one depends on its dimensions. In the following sections we will describe both issues and their consequences for the application of the MTM method to protein systems and provide efficient solutions.

Roughness of the Potential Surface

The most confusing result is that if the roughness of the potential surface Δ is chosen to be $\Delta \gg \beta$, the balance of the MTM transition probabilities between going down or up in energy is completely shifted and transitions upwards in energy will be favored. This means, that if the potential decays too fast, MTM cannot follow its gradient. The reason for this behavior is the irrelevance of the first sampled ensemble in comparison with the second one (Figure 5.1). A simple solution is to introduce a hard limit for the energy difference ΔE, but this leads to another free parameter. Also MTM could still be slowed down unnoticedly due to potential roughness. A much smarter solution is to introduce a $\lambda \neq 1$ which eliminates the diverging exponential functions. The only possible solution of choosing λ for eliminating all exponential terms is:

$$\lambda(\vec{x}_t, \vec{y}) = \frac{1}{\pi(\vec{x}_t) + \pi(\vec{y}))} \quad . \tag{5.4}$$

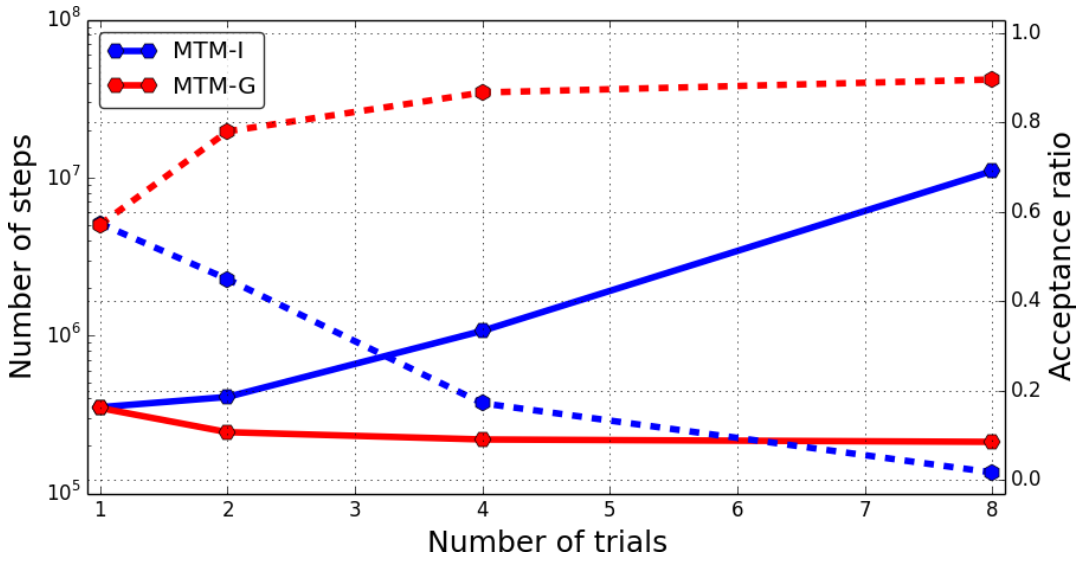


Figure 5.2.: *Minimizing a N-dimensional parabola function* $U = \sum_{i=1}^{N} x_i^2$*, the newly chosen* λ*-function clarifies its benefits. Comparing the number of required steps, to reach the minimum of the potential energy of a simple MH, a MTM-I and the modified MTM-G, reveals the MTMs capabilities. However, the benefits are limited by the acceptance ratio (dotted). We commenced each minimization at position* 30 *within all* $N = 1000$ *dimensions applying step sizes of* 0.2 *at* $kT = 0.6$*.*

It limits each $\omega(\vec{x} \to \vec{y})$ to:

$$\begin{aligned} \omega(\vec{x} \to \vec{y}) &= \frac{\pi(\vec{y}) \cdot T(\vec{x}, \vec{y})}{\pi(\vec{x}) + \pi(\vec{y})} \\ &= \frac{1}{1 + e^{-\beta(E_{\vec{x}} - E_{\vec{y}})}} \qquad \text{with } T(\vec{x}, \vec{y}) = 1 \\ &= 0.5\left(1 - \tanh\left(0.5\beta(E_{\vec{y}} - E_{\vec{x}})\right)\right) \quad , \end{aligned}$$

and thus successfully restricts it to: $0 \leq \omega(\vec{x} \to \vec{y}) \leq 1$. A further positive side effect of this choice is that this restriction also helps to reduce numerical inaccuracy. Surprisingly, the resulting weight function ω equals the Glauber acceptance criterion[212]. Hence, we will call this method the MTM-G from now on.

Applying the MTM-G method to an N-dimensional parabola potential clarifies the payoff of the chosen λ-modifier. Where heavy gradients hinder the default MTM from climbing down the potential surface, the new algorithm successfully finds the minimum of the potential, even faster than a MH method (Figure 5.2). However, enhancements coming with the modified MTM are limited by the acceptance ratio of the proposal steps and thus the number of required minimization steps decreases with increasing number of trials until the acceptance ratio reaches ≈ 1. At this point, no further computational effort will improve the minimization speed. For the common MTM the worst case scenario comes true and with increasing proposal size the sampling speed decreases; even the smallest MTM with $N = 2$ trials the MH method.

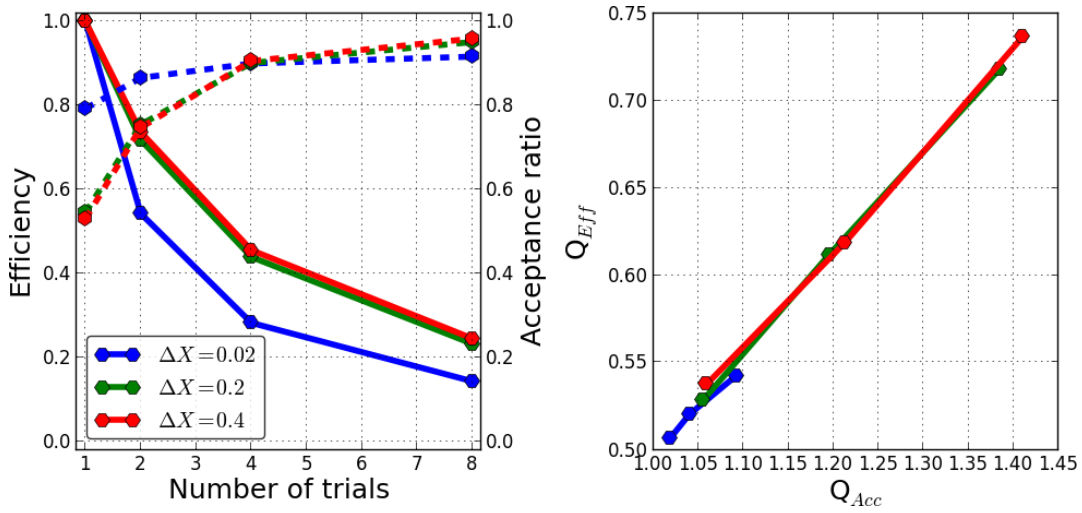


Figure 5.3.: *Minimization of the N-dimensional parabola using different steps sizes ΔX. Larger step sizes resulted in higher efficiency compared to MH simulations (left). However, the stepwise efficiency is highly correlated with the according change of the acceptance ratio (right).*

Inefficiency resulting from high Acceptance Ratios

To obtain a measurable value for the trial size depended efficiency of a MTM simulation we divide the number of steps required for a MH minimization by the number of MTM-steps as well as by the trial size N. The resulting efficiency was strongly correlated with the increment of the MH acceptance ratio (Figure 5.3). Thus, to ensure a proper efficiency of the MTM-G simulations, the complexity of the transition function has to be increased in order to achieve low MH acceptance ratios.

Generalizing the potential surface to higher dimensions, also including monotonous energy profiles, reveals yet another issue that we have to consider in order to pitch the boundaries for a well-working MTM method. Introducing too much flat dimensions, which have no influence on the potential energy, obviously increases the overall acceptance ratios of MH and MTM simulations. By introducing a certain amount of zero-transitions to the transition function, the MH acceptance ratio could be artificially increased and also simultaneously the duration of the minimization. Although having high acceptance ratios of nearly 100%, the MTM-G method still can achieve appropriate enhancements by mainly traversing through non-monotonous dimensions. Yet we can observe a slightly decrease of the efficiency (Figure 5.4).

5.3. MTM for Proteins

After having understood the MTM method by exploring its range of validity in a simple low dimensional potential system, we try to analogize the explored boundaries onto the many dimensional and rough potential energy landscape of a protein. Eliminating both difficulties, rough potential surfaces and diffusive dimensions should result in a robust and parallel alternative to the MH algorithm.

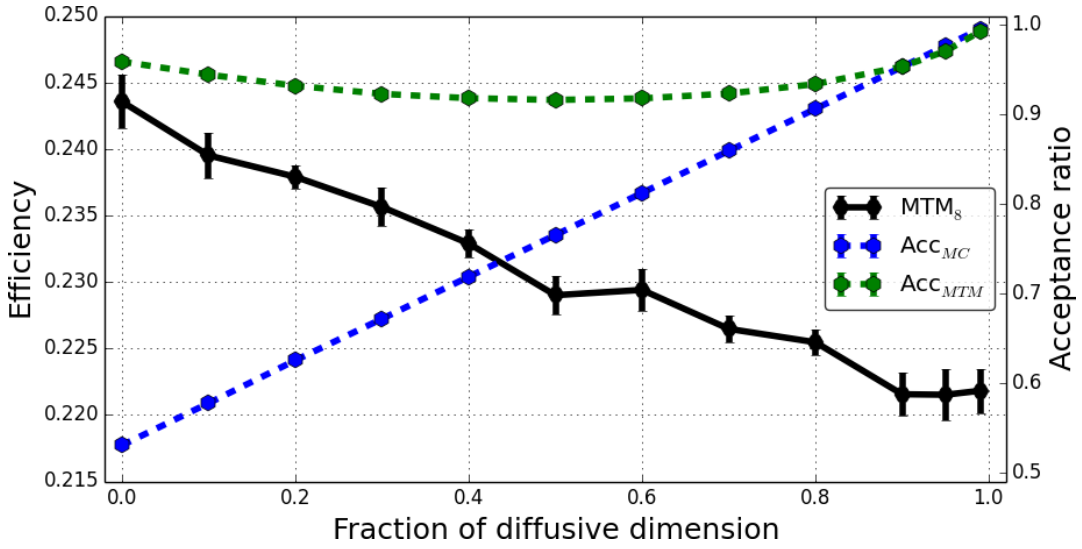


Figure 5.4.: *Despite high acceptance ratios, all sampling techniques decrease propagation through interesting dimensions after a certain amount of energetically monotonous dimensions have been introduced. However, we could only observe a slightly loss of efficiency for the MTM-G compared to MH.*

5.3.1. Energy Landscape Roughness

The complex potential energy function of a protein has been theorized to be a rough funnel. When starting the simulation from initially unfolded states, large gradients can be expected due to the initial hydrophobic collapse. Likewise on the N-dimensional parabola, the unmodified MTM-I method will not be able to accept any found state in these regions and thus would be useless. Although these special regions of the proteins potential surface are statistically completely unimportant and will not be ever populated, other large energy differences may appear somewhere on the energy landscape and could slow down the sampling speed. A behavior, which can lead to slower exploration speeds compared to an MH approach.

In protein folding simulations using the unmodified MTM-I, we could exactly observe the expected behavior of never accepting any proposed state, whereas a common MH approach virtually causes extended peptide chains to collapse instantaneously into compact structures. The introduction of the earlier proposed $\lambda(\vec{x}, \vec{y})$ function from equation 5.4, eliminates this problem and permits the MTM-G method to also collapse extended peptide chains. With the more flexible generalization to the MTM method, the GMTM, we can now freely change our condensed selection properties:

$$\omega(\vec{x} \to \vec{y}) = 0.5\left(1 - \tanh\left(0.5\beta(E_{\vec{y}} - E_{\vec{x}})\right)\right) \quad ,$$

by modifying the amplitude A and inclination B of the function:

$$\omega(\vec{x} \to \vec{y}) = A\left(1 - \tanh\left(B\beta(E_{\vec{y}} - E_{\vec{x}})\right)\right) \quad . \tag{5.5}$$

Fine-tuning both parameters can give yet another boost in efficiency (Section 5.3.3).

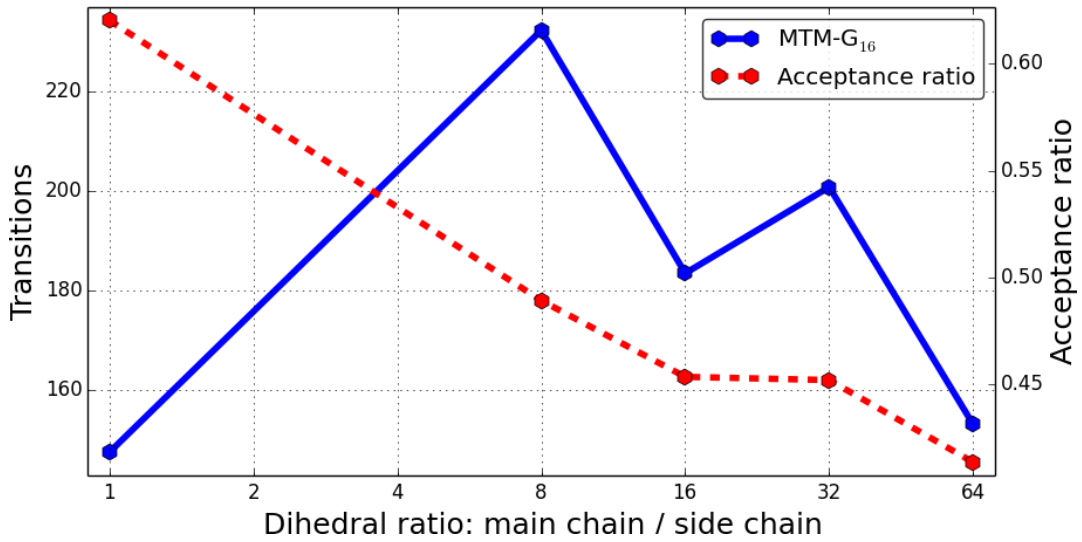
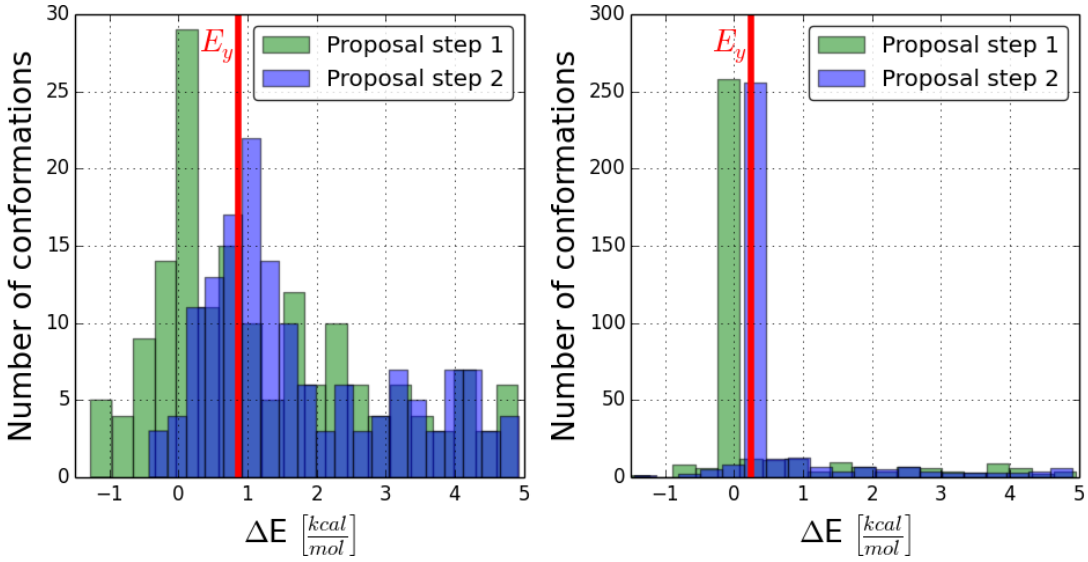


Figure 5.5.: *SIMONA is able to adjust the weight of every single transformation. An importance weighting to perturb main chain dihedrals more likely than side chain dihedrals also increases the transition frequency. However, this effect is rather owed to the effective increment of the step range of the transition function than to the changes on the overall acceptance ratio.*

5.3.2. Inefficiency resulting from high Acceptance Ratios

The second problem of slowing down the MTMs exploration speed through diffusive motions within energetically monotonous dimensions seems more tricky. However, those dimensions correspond to single dihedral angles within the peptide chain. In general, only side chain dihedral angles may have minor influences on the potential energy, whereas main chain dihedrals strongly affect the global conformation. By weighting single transformations separately, SIMONA is able to perform main chain dihedral rotations more likely than side chain rotations, however, the consequent change of the acceptance ratio only scarcely affects the sampling efficiency (Figure 5.5). We think that the observed benefits of the dihedral weighting are rather accounted for by an effective magnification of the step range within the transition function $T(\vec{x} \to \cdot)$. Thus, we could not observe an equivalent behavior for this issue.
On the contrary, in simulations where the transition function was represented equally weighted by either a dihedral rotation or a backbone rebuilding move, we could observe a drop of the efficiency. By analyzing the energy distributions of both MTM proposal steps, we found highly populated energy sectors around the respective starting energy (Figure 5.6). As the backbone rebuilding move has a low success ratio, this scenario equals the introduction of 50% flat dimensions in the model system, which cause the problem of decreasing the sampling efficiency.

The optimal choice of the transition kernel allows us to perform efficient MTM simulations for proteins. For exploiting the whole parallelization capacity, this transition kernel should rather include most complex transitions with low acceptance probability than simple but more likely to accept transformations. If the overall acceptance ratio is low enough, the trial size N can be increases until this ratio has reached an acceptable value and the best parallelization efficiency can be achieved. However, MH simulations may exhibit large acceptance ratios and nevertheless an MTM approach could effectively enhance the sampling if those were caused by transitions within diffusive dimensions.



(a) *Dihedral only.* **(b)** *Additional loop rebuilding move.*

Figure 5.6.: *Analog to the introduction of monotonous dimensions in the model system, adding transitions which do not influence neither the configuration nor the potential energy, resulted in freezing proteins. Consequently, the peptide chain was no longer able to adopt any further conformational changes.*

5.3.3. Tuning the Efficiency

To compare the efficiency of two distinct protein simulations, we count the number of complete transitions between the folded and the unfolded state within a reaction coordinate Q. The choice of an adequate Q, in the context of protein folding, is explained in chapter 6.2.3. To rate the efficiency of a simulation method it is sufficient to know that Q projects a proteins conformation onto a one-dimensional, two-state system. The number of barrier crossings between those two states, namely the folded and the unfolded state, gives us a measurable quantity for the exploration speed through the phase space of the protein.

In chapter 5.2.3 we observed that the efficiency of the MTM method is highly correlated with acceptance ratio of the proposed trial conformations. Analog to the model system, the influence of a single transition to the potential energy varies. In the proteins dihedral space this can be illustrated by sequential position, secondary structure and type of each dihedral rotation. Figure 6.9 shows the acceptance ratios for a MH simulation of the Villin headpiece[213]. Here, rotations of side chain dihedrals were accepted twice to triple as often compared to main chain dihedral rotations. The acceptance ratio is also higher for transformations within the first and the last part of the primary sequence and lower for buried parts within helical regions. However, the most surprising observation was the overall high acceptance probability for the underlying Gaussian distribution with $\sigma = 20°$. In order to improve the MTM sampling efficiency by lowering the acceptance ratio, we abandon the Gaussian distribution for the rotation angles, which has turned to account in MH simulations, and replace it by variably broadened equal distributions. As shown in figure 5.7, this change has dramatic influences on the transition rates. For a fixed MTM-G size of $N = 32$ the simulations perform up to 9 times more transitions by increasing the width of the equal distribution.

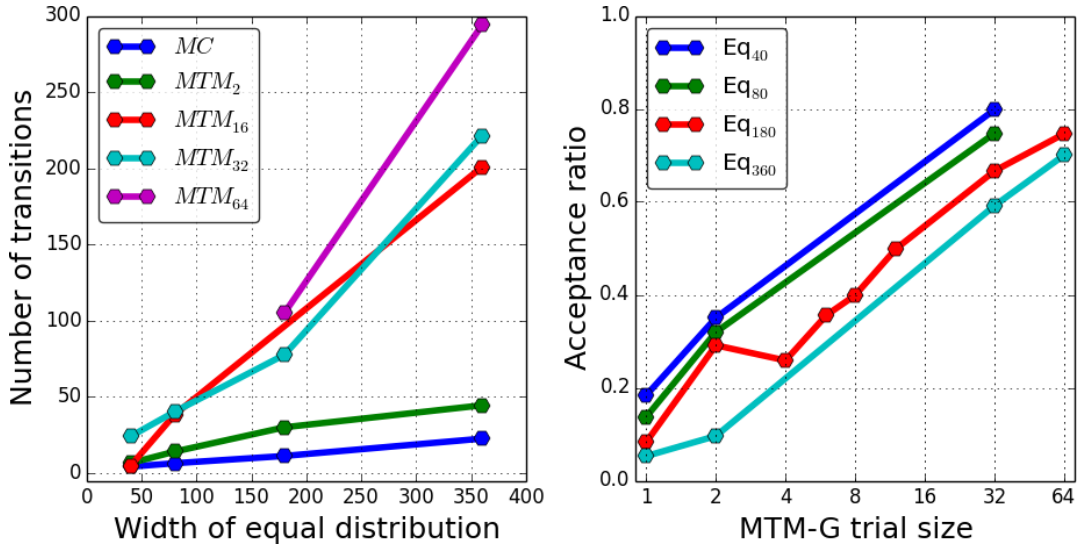


Figure 5.7.: *Influences of the chosen dihedral distribution on the transition frequency between the native and the unfolded conformation. The width of the equal distribution strongly influences the transition counts per* $100M$ *simulations steps (left). The acceptance ratio for each distribution grows with increasing MTM trial size N (right).*

Another way for determining the efficiency of a sampling method is the calculation of the auto-correlation function. The auto-correlation function for the reaction coordinate Q defines the similarity of two configurations as function of steps (lags) between them. The faster this signal drops to zero, the more efficient is the underlying sampling method. The analyses of our protein simulations pointed out that with increasing trial size N the MTM-G approach becomes a more efficient sampler (Figure 5.8). Again, we could observe that with more complex transition functions the MTM can achieve better speedups in comparison to MH simulations (Figure 5.9).

After we could successfully apply the MTM-G method to protein systems by finding the optimal transition kernel and thus observe reversible folding events, we now try to find the optimal settings for both arbitrary parameters: A and B in the weight function: ω (Eq. 5.5). An MTM-G method having modified either one of these two parameters will be called MTM-G*. For comparison we performed nine simulations with $A, B \in \{0.25, 0.5, 1.0\} \otimes \{0.25, 0.5, 1.0\}$. All simulations show various folding / unfolding transitions, however, the choice of $A = 0.25$ and $B = 0.5$ seems to be the most appropriate selection (Figure 5.10.

So far, we could demonstrate that the extension of the common MTM method to the MTM-G* method, presented here, can be applied to protein systems with an increase of the sampling efficiency compared to conventional MH simulations. The efficiency was assessed by a protein simulation's rate limiting step of crossing the free energy barrier between the folded and the unfolded state. However, we determined the sampling efficiency based on only the total number of steps. For obtaining the real efficiency, we have to further include the total arising computational effort. To compare the MTM-G with the MH in terms of CPU-hours, we have to divide the number of MTM transitions by a factor of two, to include the second proposal step, by the number of trials N and finally by a factor of ≈ 1.5 to take into account that the parallel energy evaluation of multiple conformations is limited by the slowest one. Although the MTM-G produces unseen

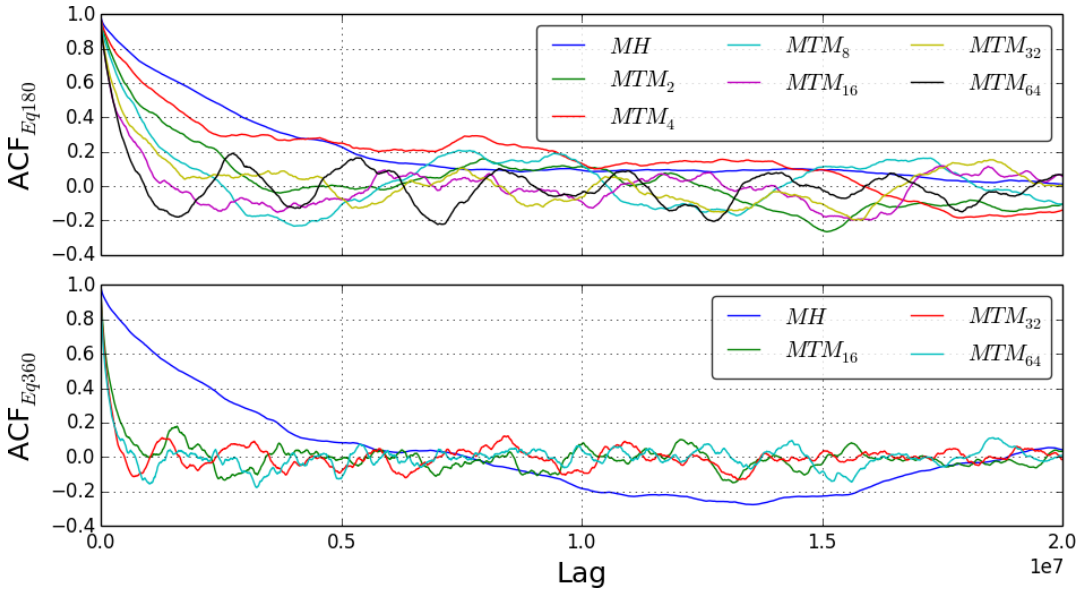


Figure 5.8.: *The auto-correlation function defines the similarity between to configurations as a function of steps (lags) between them. The higher the trial size N the faster the signal decays, implying a more efficient sampling method. However, for larger widths of the underlying perturbation distribution* $180°$ *(top) and* $360°$ *(bottom), the enhancements are more pronounced.*

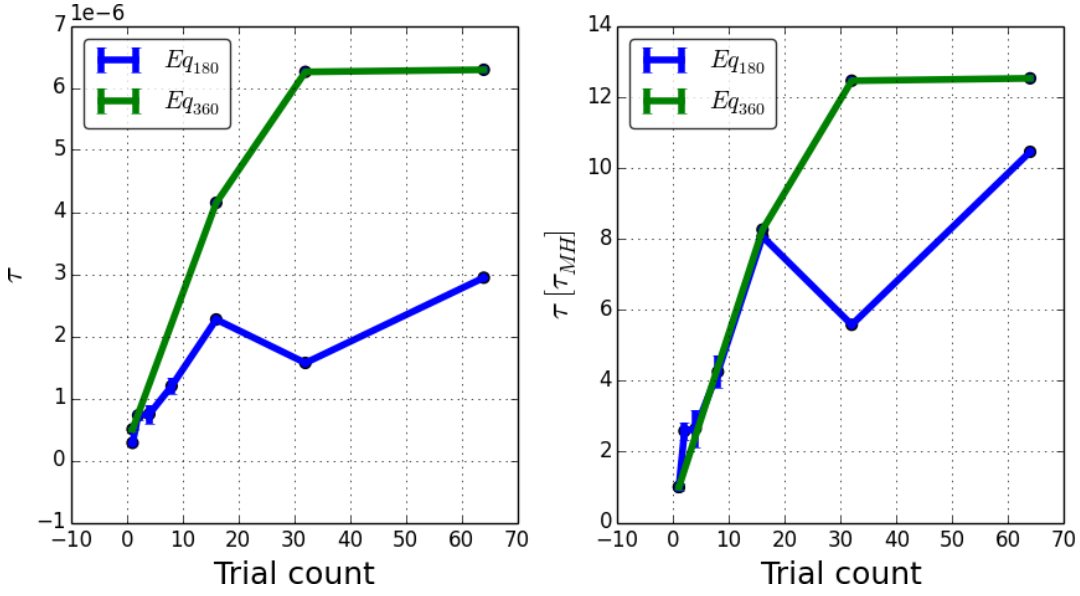


Figure 5.9.: *The decay constant for the auto-correlation functions (Figure 5.8) were obtained by fitting an exponential function to each signal. Again, we can observe better enhancement capabilities for broader chosen transition distributions. The MH data point resides at a trial count of one.*

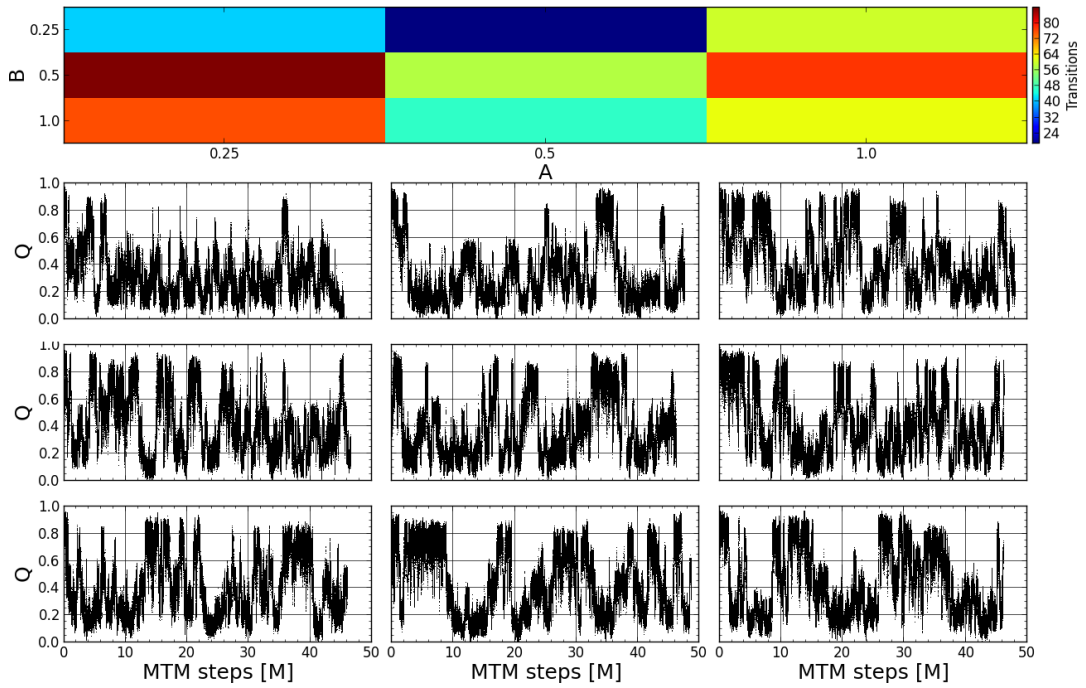


Figure 5.10.: *Fine-tuning the MTM-G to optimize the number of folding transitions per step. The figure shows the number of transitions of the Villin headpiece subdomain between the folded ($Q \geq 0.75$) and the unfolded state ($Q \leq 0.2$) normalized to $100Mio$ steps depending on both parameters A and B of $\omega(\vec{x} \to \vec{y}) = A\,(1 - \tanh\,(B\beta(E_{\vec{y}} - E_{\vec{x}})))$. Setting $A = 0.25$ and $B = 0.50$ seems to be a good choice, however, no clear trend is observable.*

trajectories including many reversible folding events within much fewer simulation steps than a MH simulation, with the current transition kernel comprising only a single dihedral angle, the computational efficiency of an MTM-G approach lies at 30% (Figure 5.11).

However, compared to an MD simulation, the required computational effort for a MTM-G simulations is rather low. One core of the Anton supercomputer comprises 512 identical application-specific integrated circuits (ASIC), optimized for MD calculations[214] and should grant a speedup of $9,000$ compared to a single conventional processor[215]. If we now compare the force field evaluation times with a common x86 CPU, one of these cores corresponds to approximately $2,500$ x86 cores[216]. The explicit water simulations of the Villin headpiece involved $\approx 5,000$ atoms, which should result in a simulations speed of $20\mu s$ per day[112, 216] and thus, a $100\mu s$ trajectory would require $300,000$ CPUh. In contrary, our implicit MTM-G_{32} simulations required: $32\text{cores} \cdot 4\text{threads} \cdot 100\mu s/17\frac{\mu s}{day} = 18,000$ CPUh what makes them 17 times more efficient[1].

The next logical step for increasing the efficiency gain by lowering the acceptance ratio of a single MTM-G step would be to perform multiple dihedral rotations at once. Besides the efficiency concerns we think that a simultaneous rotation of all dihedral angles could result in a more MD like behavior by traversing on the integral gradient of the energy landscape of a protein. A further way to increase the efficiency up to a factor of two would be to reuse the generated ensemble of the second trial generation step for the first one of the successive MTM

[1] The time-equivalent efficiency of $17\frac{\mu s}{day}$ per MTM-G step was approximated by experimental folding time of $8\mu s$ and the number of folding events in the simulation.

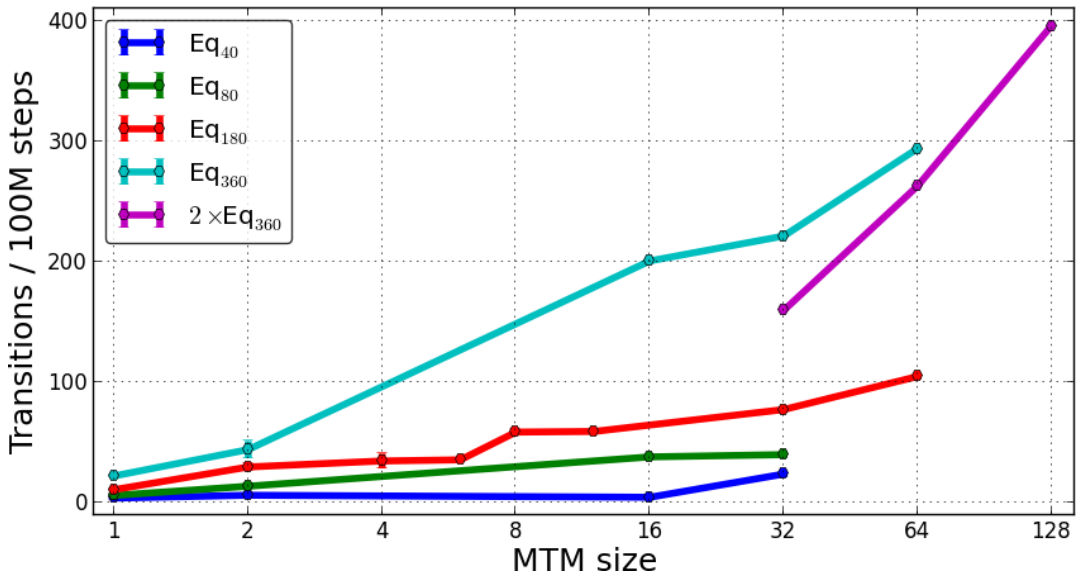


Figure 5.11.: *Influences of the chosen dihedral distribution on the transition frequency between the native and the unfolded conformation. The width of the equal distribution strongly influences the transition counts. Using more complex transition functions, the trial size of the MTM-G method can be increased to achieve further speedups. Altering multiple dihedral angles per MTM-G steps seems to be a promising way for performing massively parallel simulations with high trial sizes.*

step. However, this could only be done if the proposed conformation was accepted and thus the enhancement accounts approximately two times the acceptance ratio.

5.4. Conclusions

The Multiple Try Metropolis method was proposed 13 ago and successfully applied to theoretical potential functions. By using the same amount of computational effort, shorter autocorrelation times could be measured compared to MH simulations, which indicates a faster phase space exploration. However, the MTM method was never successfully applied in the field of long time scale molecular simulations, one of which is the protein folding process. This could be related to the lack of appropriate Metropolis based simulation frameworks for proteins, to the enormous computational effort that must be raised for these simulations or to the difficulties, which arise by increasing the complexity and dimensionality of the examined problem.
Our extensions to the conventional MTM method created a robust and fast MTM-G method, by which we could successfully improve the description of highly dimensional systems exhibiting a rough-potential surface like protein systems. The right choice of the arbitrary parameters significantly increases the phase space exploration speed and results in an improved thermodynamic description of a protein in comparison to Metropolis Hastings simulations.
Results for the successful application of the MTM-G method to the folding of the natural Villin headpiece protein are given in chapter 6.3, also including a direct comparison between the MTM-G and both established simulations techniques MD and MH. With promising parallelization capabilities we could successfully perform reversible folding simulations of the Villin headpiece subdomain, which exhibit multiple folding and unfolding events. By observing two folding / unfolding transition per day, our MTM-G approach facilitates simulations of the protein folding

process within a single day.

6. Protein Folding Simulations

6.1. Motivation

The folding process of a protein describes its transition from an unstructured chain of amino acids to a distinct three-dimensional conformation. This process is so complex that a random search for the native structure would take longer than the lifetime of the universe. In the current view, the landscape of a protein's potential energy is shaped like a funnel: a gradient towards the native structure. Due to the high dimensionality and roughness of this funnel, a minimization of the protein's energy function to locate the minimum and thereby predict the 3D-structure is very complicated. But the knowledge of the structure is crucial for understanding the functionality of the protein. Even the slightest disturbance to the native conformation of a protein can change its function. Protein systems exist, which fold into completely different motifs upon mutation of only a single residue. The extra-cellular domain of the protein G, found in streptococci, can fold either into a three helix bundle (G_A conformation) or, upon mutation of a single residue, into a four-strand β-sheet G_B motif including a single α-helix[217]. Similarly, the β-amyloid $39-43$ protein ($A\beta$) aggregates to macroscopic amyloid fibers implicated in Alzheimer's disease, while its truncated relative, containing 38 residues, cannot[218]. Thus, simulations of the protein folding process require both, speed due to the large time scale involved and a high simulation precision. In the past we were able to predict the native structure of smaller sized proteins by using evolutionary algorithms implemented in the simulation package POEM[130, 219].

Although the tertiary structure could be predicted within experimental resolution, it was impossible to adequately sample the configurational space to generate thermodynamic ensembles containing multiple folding and unfolding events. Due to the rapid development in computational science, involving the enhancement of force fields and the advance of supercomputers, the theoretical description of proteins increasingly becomes more and more important. Recent investigations demonstrated that MD simulations are now able to simulate proteins up to the millisecond time scale within a physical, atomistic force field[95, 112, 113]. In spite of the impact of the fast simulation techniques developed by Shaw et. al[95, 214, 216], the methods are not readily available to the general public as sufficiently long MD simulations require immense computational efforts[57] and currently only allow investigation of fast folding proteins with folding times below 100 microseconds. At the same time the theoretically achievable parallelization speedup is limited by the serial nature of most MD sampling techniques.

In order to be able to simulate large conformational change in biomolecules, we have developed the efficient and versatile simulation framework SIMONA. In addition to our in house force fields PFF01/02 we also enabled the parametrization of a protein within the AMBER99SB force field, a popular force field in the MD community that correctly describes many aspects in folding

processes. Using the new simulation package we could perform long MH simulations of proteins and in addition, we were able to implement massively parallel Metropolis methods for an even faster thermodynamic characterization of proteins.

6.2. Methods

6.2.1. AMBER Force Field

The development of the AMBER[109, 124–126] force field spans the last two decades. It comprises various potential functions derived from experimental observations and was fine tuned to correctly describe many aspects in the behavior of proteins[127, 128, 220]. The latest modification, the AMBER99SB-STAR-ILDN force field, was shown to be ideally suited to describe the folding of both helical and β-sheet topologies and was used recently for a complex study of mutations of the Villin headpiece subdomain[221].

6.2.2. PFF03 Force Field

Based on the AMBER99SB-STAR-ILDN parameter set, including its Lennard Jones parameters, partial charges and torsion constraints, we developed the PFF03 force field. As the explicit handling of solvent molecules is difficult in Monte Carlo based simulation techniques, we use an implicit description of the environment. Because of steric clashes between the simulated protein and the solvent molecules in explicit water simulations, a Monte Carlo method would never be able to generate physical proposal configurations and thus the acceptance ratio would drop to zero. The implicit treatment of the protein-solvent interactions is approximated by an effective GB solver with high accuracy[77].

The PFF03 force field comprises the following potential terms:

- **Lennard Jones** potential for modeling the Pauli repulsion as well as the dipole-induced London attraction.
- **Coulomb electrostatics** for calculation of partial charge interactions.
- **Non-polar sasa energy** defines the non-polar part of the solvation free energy.
- **GB solvation energy** computes the polar part of the solvation free energy.
- **Dihedral potential** applies the AMBER specific dihedral constraints to the protein.
- **PowerBorn** calculates the Born radii of each atom.
- **PowerSasa** determines the solvent accessible surface area.

For computational optimization the evaluation of all non-bonded interactions is grouped into a single potential within the SIMONA code. This allows us a more efficient parallelization on shared memory architectures. Depending on the size of the simulated protein, speedups of up

to a factor of eight can be achieved for the energy evaluation times of this force field (Figure 3.10).

6.2.3. Reaction Coordinate

The choice of a proper reaction coordinate is crucial for the analysis of protein trajectories. For the comparison of identical protein structures, the RMSD value is an appropriate method. The value of the root mean square deviation of the positions of all atoms provides information about how related two structures are to each other. The RMSD to the native structure is defined as:

$$RMSD = \min\left(\sqrt{\frac{\sum_i^N (\vec{r}_{0i} - \vec{r}_i)^2}{N}}\right) \quad ,$$

where $\vec{r}_{0i}$ represents the coordinates of the i-th atom within the native configuration and $\vec{r}_i$ the according atom, usually in a frame of a trajectory. Prior to the evaluation of the distances, both structures have to be aligned to each other, denoted with the $\min()$ function.

However, the use of the RMSD as a reaction coordinate is not always a good choice, as small changes of a protein's structure can involve huge changes in the RMSD value. The change of a single dihedral angle in a helix segment can entail the wrong classification of the structure into the unfolded state, even though most of the atomic contacts are still developed. Rotating a single dihedral angle of the helical Villin headpiece protein results in a configuration with an RMSD of 5.8Å to the unperturbed structure, which would be classified as being unfolded (Figure 6.1). However, nearly all native contacts are formed in this conformation and thus the structure should be classified to be, at least, near-native. A widely used approach to characterize the degree of "nativeness" can be made by counting the number of native contacts, derived from the native structure, which are still present in the snapshot structure. We are using a shadow map based approach to create such a contact list, based on the native protein conformation[222]. Two atoms are considered to be in contact, if their distance in the sequence is larger than three residues and their spatial distance is smaller than a cut-off $D_{cut-off} = 6$Å. A physical contact between both atoms is only made, if no other atom lies between them. Therefore, the shadow map positions a small, radial light source on both atom centers and monitors, if further atoms, represented as spheres with a radius of $r_{\text{Atom}} = 1$Å, cast a shadow onto the contact partner (Figure 4.2). Only if both atoms are not shadowed, the contact is considered as native.

Based on an MC simulation of a protein we redefine our contact list by removing all contacts, which are either present in more than 80% or only in less than 20% of all occurring structures. The resulting contact information and the neglected contacts for the Villin headpiece subdomain are illustrated in Figure 6.2a. The redefinition of the contact map leads to a more precise classification of protein structures and results in a convenient reaction coordinate Q (Figure 6.2b). Both contact maps feature pronounced contacts along the diagonal axis. With a sequential distance of 4 residues, these contacts mark helical $1-4$-interactions and are therefore typical for α-helical protein regions.

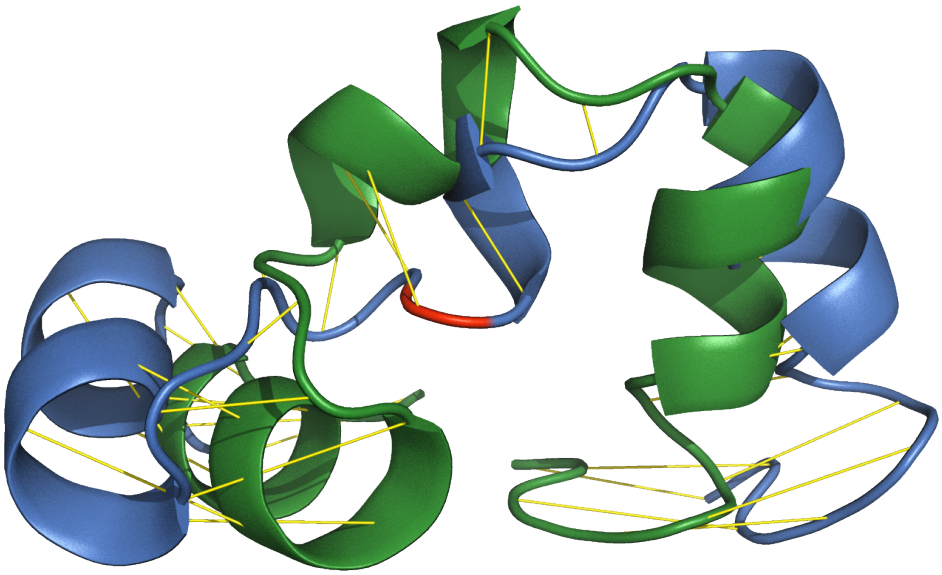
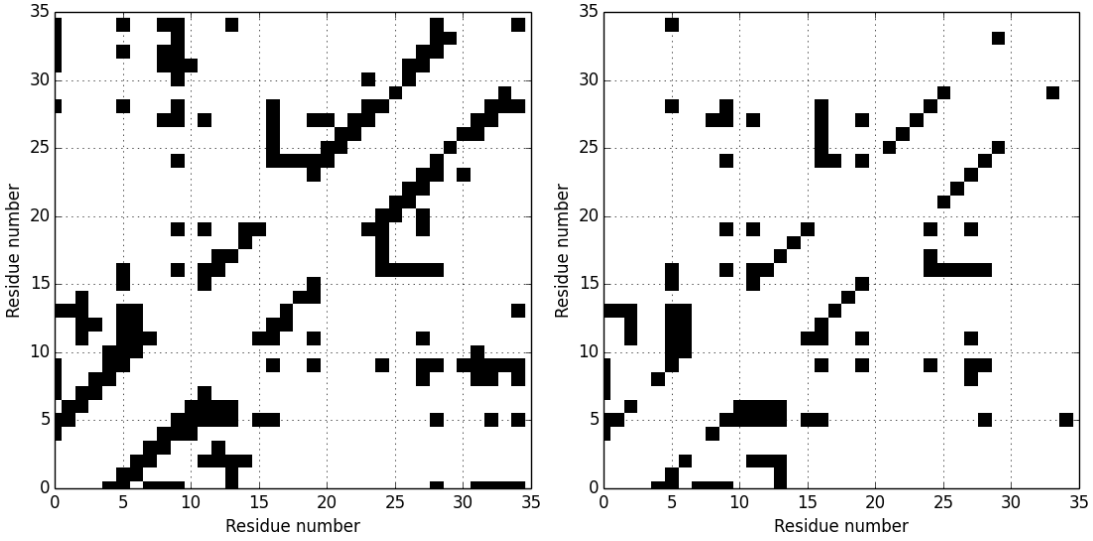


Figure 6.1.: *The figure shows the native state of the Villin headpiece subdomain (green) and a slightly modified configuration (blue) generated by only rotating one dihedral angle (red). After aligning both structures (yellow), a RMSD of* 5.8Å *can be measured, which results in a classification into the unfolded ensemble, even though the structure is almost native. This example shows the importance of a proper choice for the reaction coordinate and clarifies the issues regarding the RMSD measure.*



(a) *Original contact map.* **(b)** *Redefined contact map.*

Figure 6.2.: *The shadow map approach for the Villin headpiece subdomain generates* 815 *contacts (6.2a). After redefining the contact map by using an ergodic folding trajectory,* 334 *contacts remain (6.2b). The redefinition of the contact map allows a clearer classification of a structure and shifts the unfolded state to zero and the folded state to one.*

6.2.4. Circular Dichroism

Circular dichroism spectroscopy (CD) is a useful technique for experimentally determining the secondary structure of a protein. Even, before the first protein structure was resolved via X-ray crystallography, CD spectra could reveal the right-handed chirality of an α-helix[223–225]. A CD spectrum shows the excitation energies of the electronic structure of a protein upon an incoming circularly polarized beam of light in its characteristic bands. Every secondary structure element corresponds to a special fingerprint-like CD spectrum. Where each one features characteristic bands which correspond to special electronic transitions[226]. The comparison with reference spectra allows the determination of helical, sheet, and coil content in the investigated structure.
Even for simulations, a calculated CD spectrum can grant a deeper insight into the folding process of a protein and its optical properties[227]. A calculated CD spectrum returns the ellipticity of the structure for each wavelength which can afterwards be fitted to reference data to obtain the structural contents of the conformation[228].

6.2.5. Parallel Tempering

The Parallel Tempering method (PT) allows the structural exchange between two MH simulations, even if their temperatures are different[116, 117]. Thus, a single PT simulation can comprise a whole ensemble of temperatures of which each one corresponds to an independent MH simulation. After a certain amount of steps, simulations u, v with adjacent temperatures can exchange their conformations according to a modified MH acceptance criterion (Figure 3.8):

$$\alpha_{PT} = \min\left(1, e^{(\beta_v - \beta_u)(E_v - E_u)}\right) \quad .$$

For an optimal exchange ratio, the temperature range should be distributed exponentially and include low temperatures to freeze low energy conformations as well as high temperatures to jump over every existing barrier in the system. This additional exchange step helps to escape local minima and permits jumps over large barriers and thus results in a faster characterization of the phase space. Additionally every single MH simulation can run as a separate process and only requires sparse global communication for the structural exchange step. This makes the PT method a good target for MPI parallelization approaches.

6.3. Results

The presented results for all protein folding simulations are divided according to the simulation techniques used. We first present results from a common Metropolis Hastings approach with minimalist parallelization potential, we proceed to the Parallel Tempering simulation technique which allows the execution of multiple interchanging simulations in parallel for improving the sampling speed and conclude with our modified and massively parallel Multiple Try Metropolis

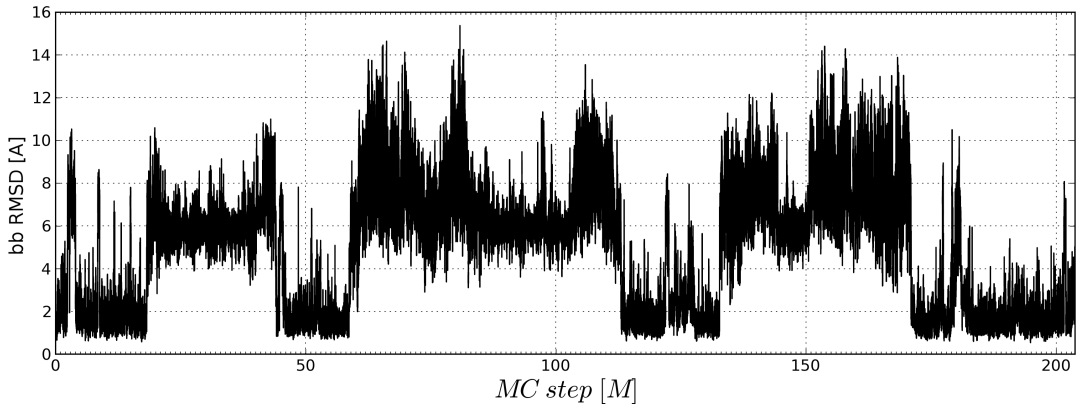


Figure 6.3.: *The simulation of the 36 residues comprising Villin headpiece subdomain (1vii) over* $2 \cdot 10^8$ *MH steps features multiple folding and unfolding events. After unfolding, associated with the complete loss of the tertiary information, the protein refolds itself within an error below* 1Å *backbone RMSD to the PDB structure. The five unstructured N-terminal residues (EKGLF) were not incorporated into the RMSD calculation.*

method.

6.3.1. Metropolis Hastings

To thermodynamically characterize the subdomain of the Villin headpiece, we carried out multiple MH simulations at the temperatures $280K$, $300K$, $320K$, $340K$, $360K$ and $380K$ with ten replicas each. In total, our simulation study comprises $8 \cdot 10^9$ MH steps with longer simulations near the folding equilibrium temperature, covering $2 \cdot 10^8$ steps per trajectory. In all simulations at temperatures high enough to allow the transition between the folded and the unfolded state, we could observe multiple folding and unfolding events. This allows us to characterize the folding barrier (Figure 6.3). In our reversible folding simulations, we could observe a complete transition from one of these two states into the other approximately every $7 \cdot 10^7$ energy evaluations. Even after the protein was completely unfolded and lost all its secondary structure information, the simulation found back again to the native configuration within an error below to 1Å of backbone RMSD compared to the experimentally resolved structure (excluding the unstructured residues $32 - 36$).
The comparison of the mean transition frequency of our simulations with experimental folding times of $8\mu s$[229] translates to a time equivalent of $116fs$ per MH step, which denotes a good simulation efficiency. If we now take into account that we can manage approximately 10^7 MH steps per day on a quad-core *Intel(R) Xeon(R) E5440* processor running at $2.83GHz$. this results in a (time equivalent) simulation speed of $1.16\mu s$ per day on off the shelf hardware, which is available to everyone.

The thermodynamic stability of the native configuration can be determined by calculating the free energy $G = H - TS$ for a given temperature. Using the complete conformational data, we could determine the free energy surface with respect to the reaction coordinate Q and obtained a one dimensional free energy profile of the Villin headpiece subdomain, which reveals three distinct states: the folded state and the unfolded state separated by an intermediate state in between of them. In order to fold, the protein has to pass the folding intermediate state, where 45%

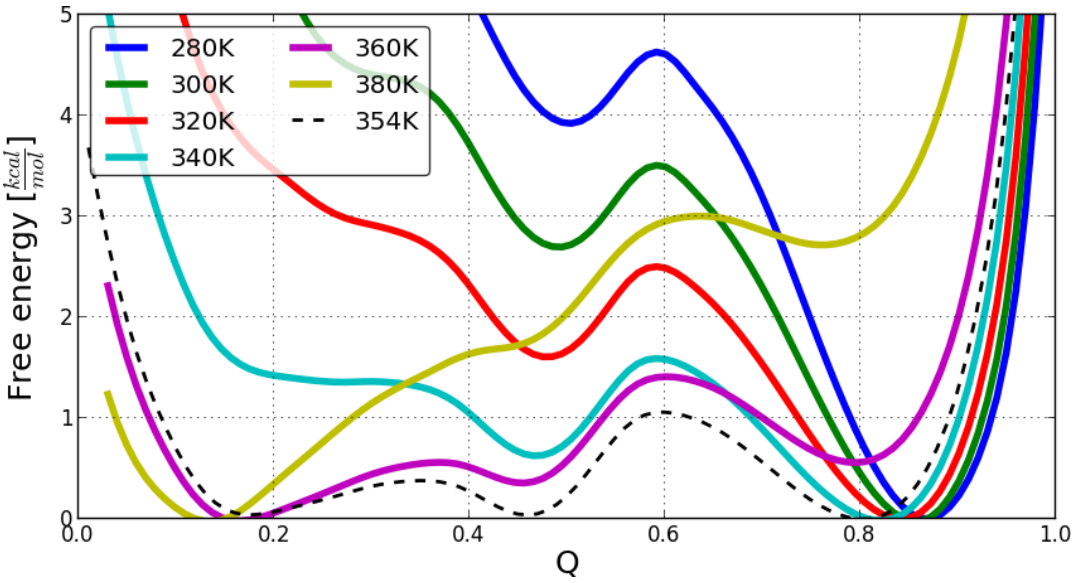


Figure 6.4.: *The free energy profile shows three distinct minima for the folded, the unfolded and an intermediate state. The interpolated free energy profile at the folding temperature of* $354K$ *shows a barrier slightly above* $1\frac{kcal}{mol}$ *between the native state and the transition state.*

of all native contacts are formed (Figure 6.4). This intermediate state was already seen in free energy analysis of prior MD studies[230]. From the free energy profile we can draw inferences about the folding equilibrium temperature of the underlying system. The folding equilibrium temperature is defined as the temperature, where both states, the folded and the unfolded, are populated equally. With our coarse-meshed temperature range we can only determine the folding equilibrium temperature to lay slightly below $360K$. We find that the interpolated free energy landscape features a barrier of approximately $1\frac{kcal}{mol}$, which stabilizes the folded state.

To link the structural information with the three state system of the free energy analysis, we investigated the per residue energy contribution and their thermal stability. Figure 6.5 reveals the native contacts for a given temperature or Q value on a per-residue basis. Both plots show a high stability of the first two helices and a slightly lower thermal stability of the third one. At the folding equilibrium temperature of $354K$, the residues within helix one and two are in a native configuration in half of the occurring structures (Figure 6.5a). However, in only 20% of these structures, residues forming the third helix are in a native conformation. This reveals a lower thermal stability of the third helix, also visible in a Q-dependent analysis, where the third helix is only in a native conformation for Q-values above 75% (Figure 6.5b). The lack of formation of tertiary contacts for residues 5, 22, $32-34$, all contact based analysis contain missing entries (A single blue bar for a residue). Due to unstructured parts for the N-terminal region in the reference crystal structure, affected residues lack tertiary contacts. Analyzing the Q-dependent contact map reveals an early helix formation in a nucleation like process (Figure 6.5c). At low Q values the contacts of the first helix dominate the contact map. This suggest that this helix is formed prior to any other tertiary contacts.
A classification of all occurring configurations into either the folded $(0.75-0.85)$, the transition $(0.55-0.65)$ or the unfolded $(0.1-0.5)$ state, permits us to calculate the simulation ϕ-values for each residue without simulating any mutants[231]. This method is less computational costly, but

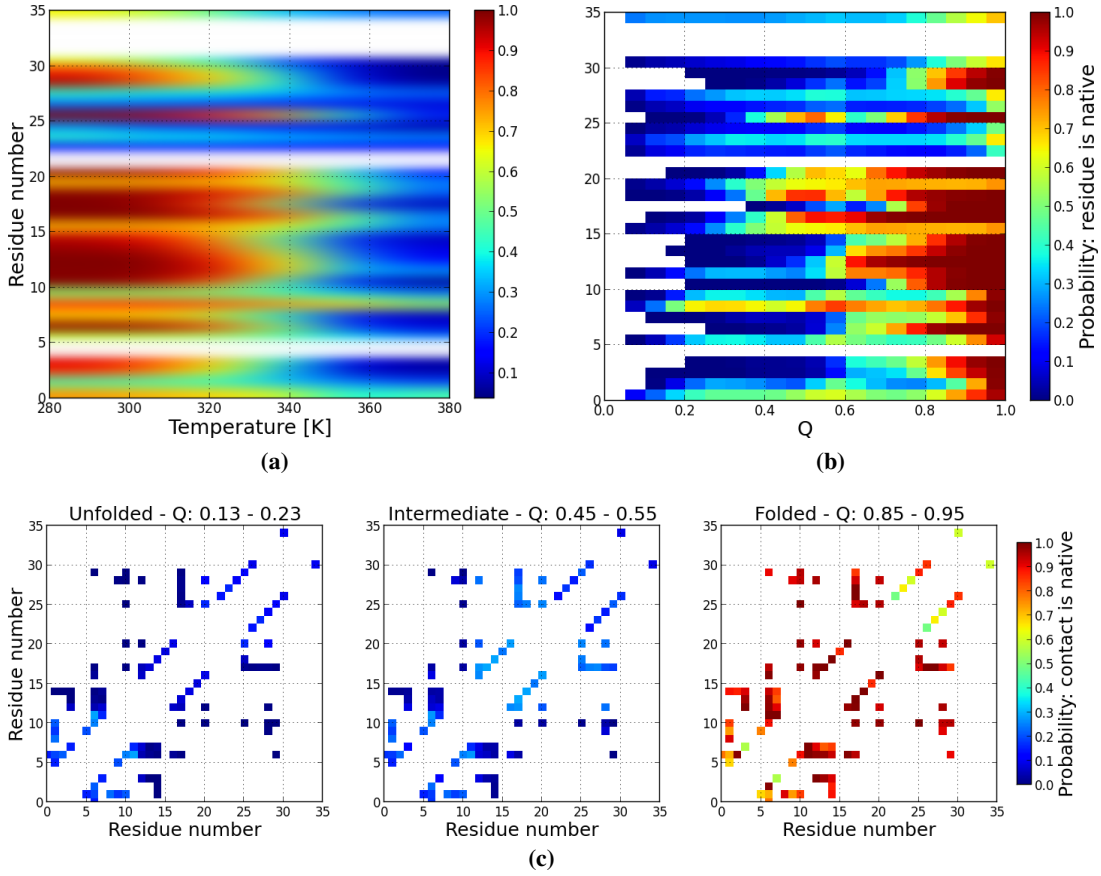
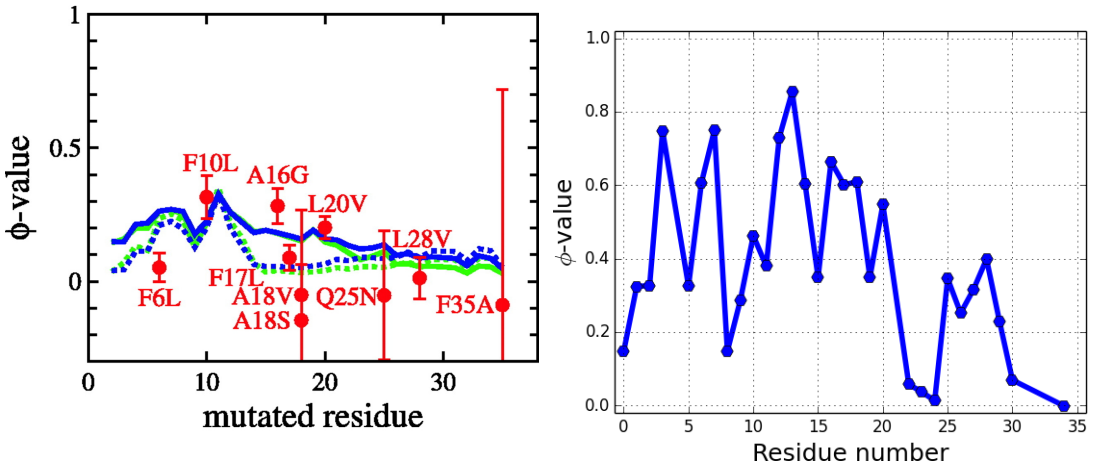


Figure 6.5.: *Each residue of the Villin headpiece subdomain has a different thermal stability (6.5a). The residues* 5*,* 23 *and* $32 - 34$ *were not involved in the reaction coordinate Q and thus no data is available. Figure 6.5b shows, which residues remain native at a given Q value. The resulting contact maps for the three distinct free energy states are shown in figure 6.5c.*

stills produces adequate results, equivalent to experimental ϕ-values, determined by perturbing the protein sequence through mutations[232] (Figure 6.6). The ϕ-value gives us a measure for the energetic contribution of a residue to the folding transition state and assigns it to either conform with the native or the non-native ensemble in the transition state. According to previous investigations, residues within the third helix have ϕ-values close to zero, which indicates an unstructured conformation in the transition state ensemble, whereas helix one and two mostly exhibit high ϕ-values and thus comprise local structures in a native-like conformation in the transition state.

To obtain more global information about the structural constitution of the simulated ensemble, we examined the secondary structure of the protein. A theoretical CD analysis can grant us more insights into the temperature stability of the secondary structure motifs and can be calculated based on a protein structure[228]. Separately for each temperature, we calculated the ensemble spectrum by averaging the CD spectra of all occurring conformations. Afterwards, we measured the ellipticity for a given temperature in the specific α-helical bands at[226, 227]:

- Positive band at $\approx 190nm$: a $\pi \rightarrow \pi^*$ transition perpendicular to the helix.
- Negative band at $\approx 210nm$: a $\pi_{nb} \rightarrow \pi^*$ transition parallel to the helix.



(a) *Experimental data; copyright © (1993–2008), National Academy of Sciences[233].*

(b) *Calculated data.*

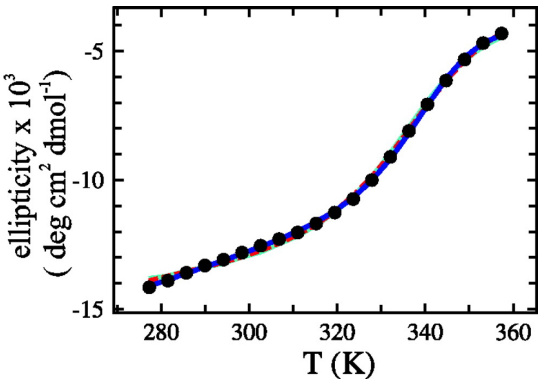
Figure 6.6.: *The fast ϕ-value analysis using a single trajectory instead of mutations for the $360K$ trajectory; (6.6b) features a similar behavior like the experimentally determined values (6.6a).*

- Negative band at $\approx 220nm$: a $n \rightarrow \pi^*$ transition perpendicular to the helix.

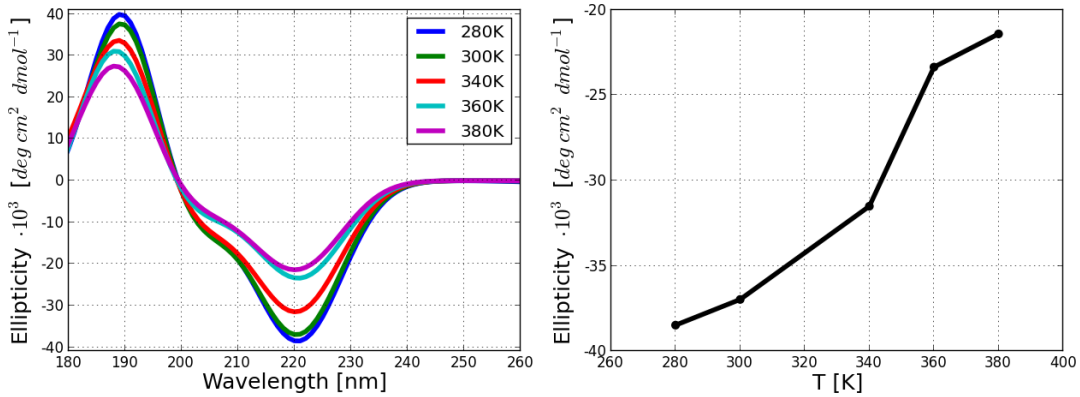
Where n is an oxygen lone pair orbital, π_{nb} an amide non-bonding π-orbital and π^* an anti-bonding π-orbital. The averaged CD spectra and the resulting temperature dependent ellipticity are shown in Figure 6.7. Obvious peaks within all helical bands suggest a mostly helical structural ensemble. As expected, the ellipticity of the ensembles decays with increasing temperatures. The temperature dependent ellipticity in the $220nm$ band (6.7b right) exhibits the same functional profile as experimental data (6.7a[233]).

It is currently unclear, why all temperatures of our implicit solvent simulations are shifted upwards. But it is still possible to rescale the temperature to the absolute temperature scale by calculating the specific heat C_v of the system and fitting it to experimental data. By the use of the weighted histogram analysis method (WHAM[234, 235]), we found an obvious peak at $T_{C_{v,max}} = 425.5K$, which yields a temperature shift of $80K$ (Figure 6.8). Thus, all Metropolis based temperatures in this thesis for simulations of the Villin headpiece were shifted backwards by $80K$ to reflect a realistic temperature. Besides the temperature shift, the $C_{v,max}$ should provide a good approximation for the folding equilibrium temperature. The temperature $T_{C_{v,max}} = 345.5K$ is located close to the earlier determined folding equilibrium temperature $T_{FE} = 354K$ calculated using the WHAM analysis.

At the moment we are investigating possible reasons and we think that either the use of an explicit MD potential within the dihedral space could account for the lacking of the energy contribution of the atomic vibrations, which may result in a temperature shift or if implicit solvent approach underestimate the Van der Waals attraction of absent water molecules. In the latter case, an extension to the common implicit GB approach was developed, which adds this missing term[236]. Analysis of these topics is complicated by the enormous computational effort of the required simulations, as only simulations on the ANTON supercomputer could reach sufficiently large time scales, to date. However, these simulations were carried out using

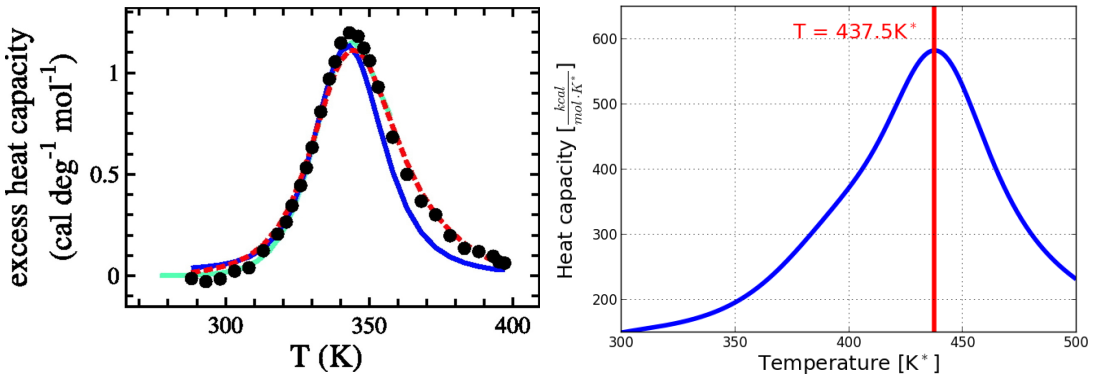


(a) *Experimental data; copyright © (1993–2008), National Academy of Sciences[233].*



(b) *Calculated data.*

Figure 6.7.: *Mean Calculated CD spectra over all occurring conformations for each temperature (6.7b left)[228]. Using the negative $\pi_{nb} \rightarrow \pi^*$ band at $210nm$ we could determine the ellipticity for a given temperature (6.7b right). The functional form matches well the experimental data (6.7a)*



(a) *Experimental ϕ-values (red) at $310K$; copyright © (1993–2008), National Academy of Sciences[233].*

(b) *Calculated WHAM data.*

Figure 6.8.: *Based on all trajectory data, we could calculate the specific heat C_v of the system using the weighted histogram analysis method (WHAM[234, 235]) (6.8b). The function is peaked at $425.5K$ and, compared to experimental data(6.8a), this indicates a temperature shift of roughly $80K$ for our simulations.*

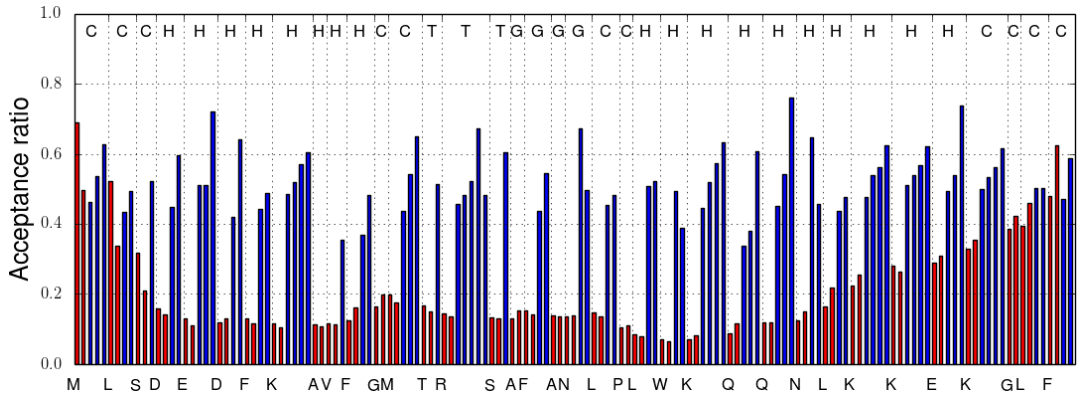


Figure 6.9.: *The per-dihedral acceptance ratio of a MH simulation strongly depends on the dihedral class (main chain or side chain dihedral), its association to secondary structure elements and the sequential position. Most side chain dihedrals were accepted with high acceptance ratios above* 40%*. In contrast, a buried main chain dihedral will be accepted in less than* 20% *attempts. In a single MH step, a random dihedral angle was perturbed randomly using a Gaussian distribution with* $\sigma = 20°$*.*

an explicit description of the aqueous environment and thus do not address these problems. Our MH simulation approach could facilitate such an investigation with high impact on related challenging problems, such as protein-protein docking, protein-ligand binding or protein folding.

The MH-transition kernel used in our simulations comprises both, random perturbations to the proteins dihedral space incorporating backbone dihedrals ϕ and ψ as well as all side chain specific sets of χ-angels, and backbone preserving loop rebuild moves. Thus, in the MH proposal step, a new trial configuration can be generated by either perturbing a single dihedral angle randomly within a Gaussian distribution ($\sigma = 20°$) or by reordering a backbone region. The acceptance ratios for the dihedral moves are astonishingly high, which would allow us to fine tune the perturbation distribution to improve the MH sampling speed (Figure 6.9). However, the acceptance ratio strongly depends on the class of the dihedral (main chain or side chain), the sequential position within the protein and the association to the protein's secondary structure. Whereas most side chain dihedrals show acceptance ratios above 40%, main chain dihedrals were only accepted in less than 20% of the steps. Simply because of the smaller influence on the global protein structure, a perturbation to a side chain angle produces clashed conformations less frequently. Besides main chain angles at the C- and N-terminal, only those within the third helix show higher acceptance ratio above 20%. According to our structural analysis, this means that this helix is less stable than both other helices. In chapter 5 we could already demonstrate that setting a dihedral angle completely random instead of only applying small perturbations admittedly decreases the overall acceptance ratio, however it increases the number of folding and unfolding transition per MH step (Figure 5.7).

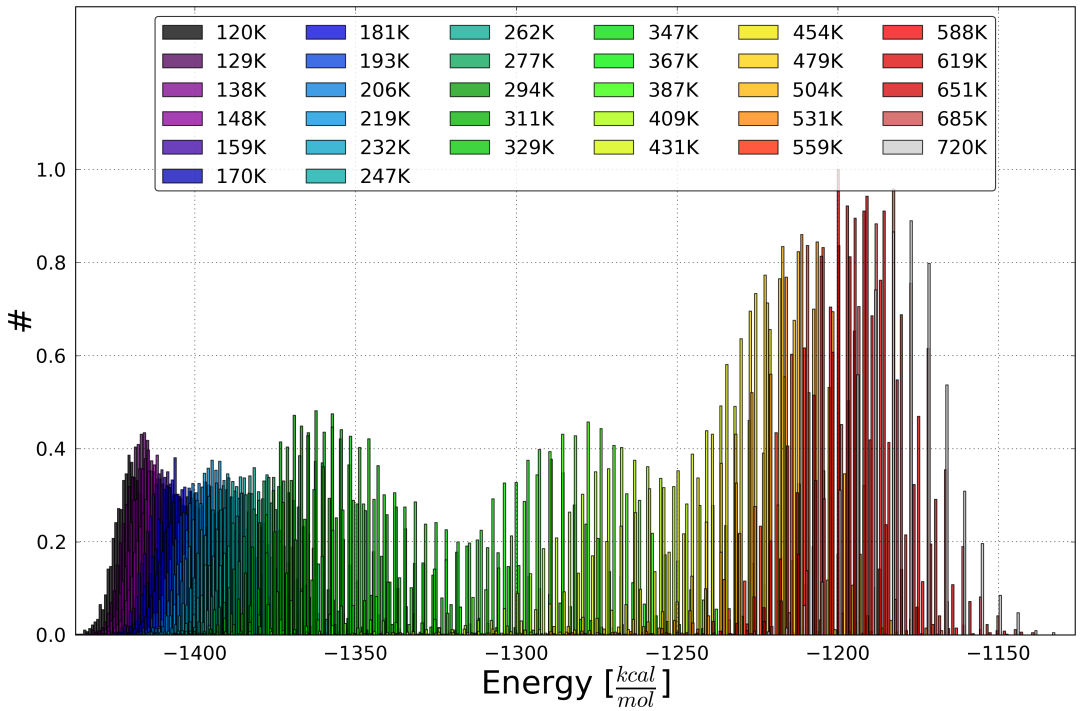


Figure 6.10.: *The PT simulation comprises* 32 *temperature basins exponentially distributed between* $120K$ *and* $720K$*. For an efficient structural exchange between two basins, both energy distributions have to overlap. Along the energy scale we see continuously overlapping energy distributions allowing exchanges.*

6.3.2. Parallel Tempering

To increase the systems phase space propagation speed, we performed PT simulations of the Villin headpiece subdomain comprising 32 temperature populations. For optimal exchange rates, the basin temperatures are distributed exponentially between $120K$ and $720K$ including low temperatures for optimizing local minima as well as high temperatures for allowing the crossing of all occurring barriers in the system. Starting from the native configuration, we performed 5000 PT steps. Each PT step comprises $10,000$ MH simulation steps per replica with subsequent structural exchanges between neighboring temperature basins. In total, the PT comprises $5 \cdot 10^7$ energy evaluations and all RMSD plots can be found in the appendix A.2.
In the analysis of the energy distributions of each temperature, we see overlapping distributions, ranging from $-1450\frac{kcal}{mol}$ to $-1150\frac{kcal}{mol}$ (Figure 6.10). Two energy gaps at $-1325\frac{kcal}{mol}$ and $-1250\frac{kcal}{mol}$ can be observed, which represent the lower populated transitions separating the intermediate state from the folded and the unfolded states. Only if we find overlapping energy distributions in this area, the PT simulation can perform barrier crossing exchanges and help to increase the sampling efficiency. With multiple temperature basins having overlapping energy contributions across transition state regions in Q space, the simulation was able to perform many barrier crossing transitions.
The ability to perform barrier crossing transitions allowed us to calculate an accurate free energy profile with much less simulation steps in every single temperature ensemble compared to conventional MH simulations. We found the same three states as before in the MH simulations (Figure 6.11). The finer temperature distribution, allowed us to determine free energy profiles

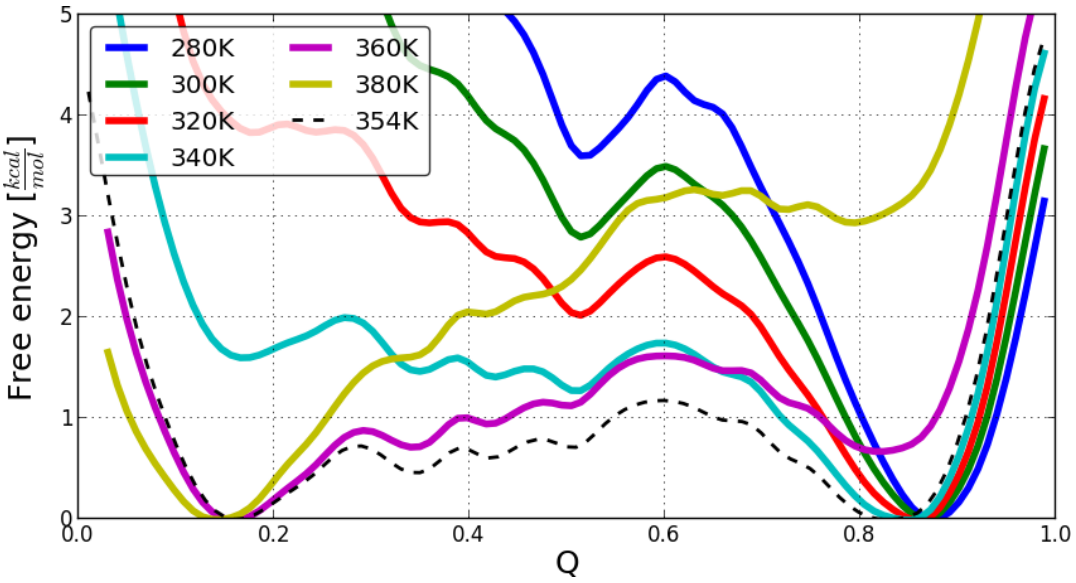
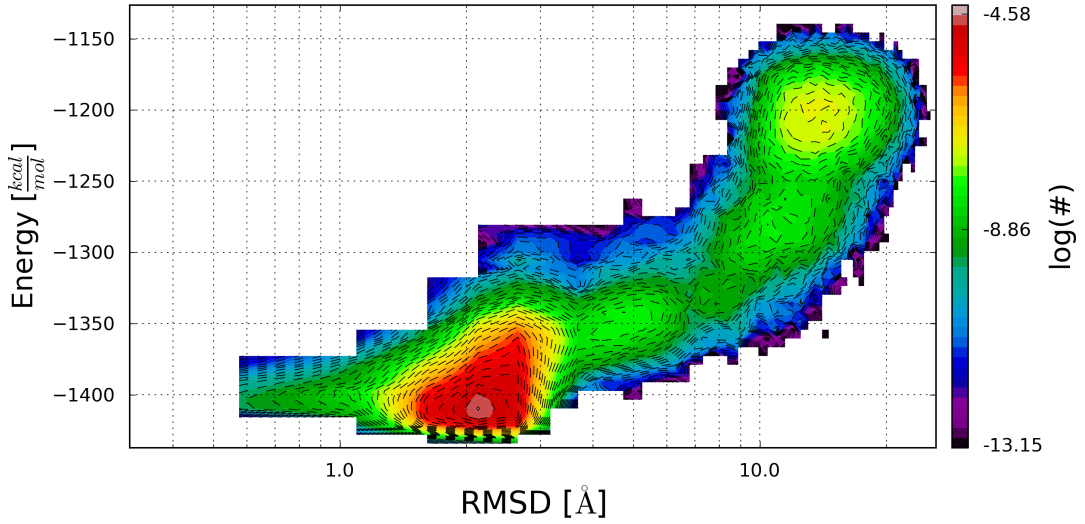


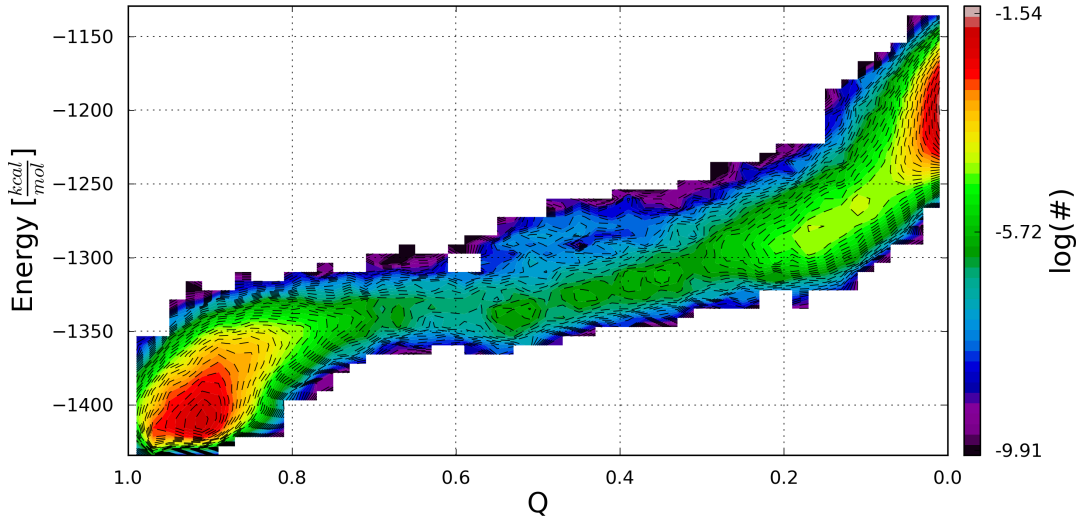
Figure 6.11.: *The free energy profile equals the previous calculation based on MH simulations, except the less complete resolved folding intermediate state at* $Q = 0.46$*. At folding equilibrium temperature of* $354K$*, the folded state is stabilized by a barrier of* $1.2\frac{kcal}{mol}$.

more accurate for a given temperature. The folding intermediate state was lower populated, but still visible. Due to overlapping energy distributions between two basins holding folded and unfolded conformations, a transition between these states could occur without traversing the intermediate state. At $354K$ the folded and the unfolded state lie on the same free energy level where the native state is stabilized by a barrier of $1.2\frac{kcal}{mol}$.

The sampled phase space can be visualized by marking each conformation in a energy vs. Q histogram (Figure 6.12). The simulation shows a good sampling an features a folding pathway between the folded and the unfolded state. However, due to the PT simulation scheme, a folding/unfolding transition can occur through structural exchanges, without following this pathway.



(a) *Energy vs. RMSD histogram.*



(b) *Energy vs. Q histogram.*

Figure 6.12.: *Both histograms of the potential energy against the RMSD (6.12a) and Q (6.12b) feature pronounced regions for the folded and the unfolded state. The lower populated intermediate state at* $Q = 0.46$ *lies in between and spatially separates both states from each other. Besides transitions traversing the intermediate state, also the structural exchange allows folding/unfolding transitions.*

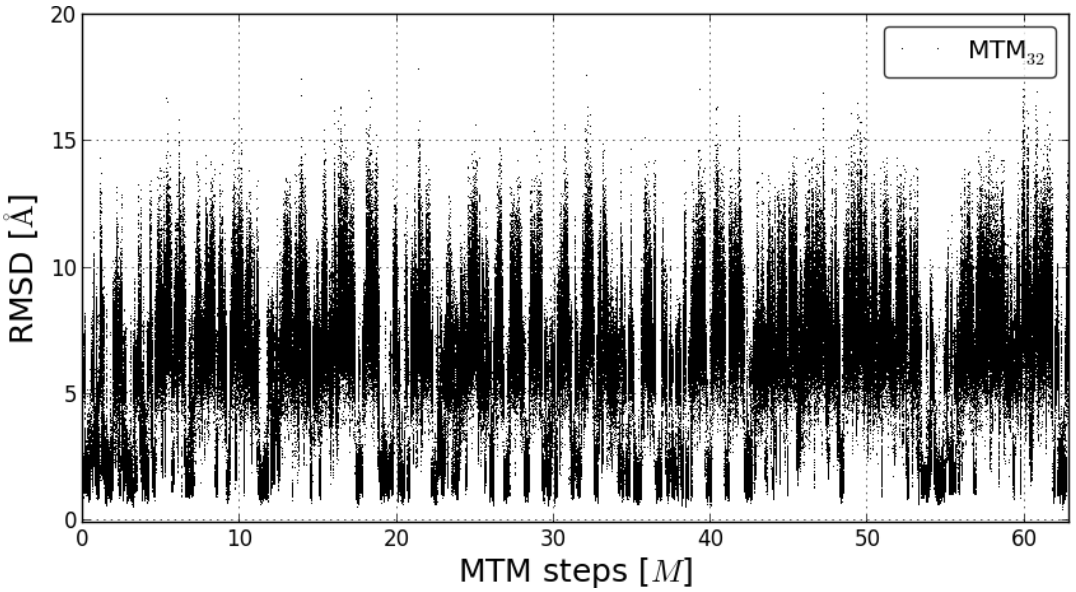


Figure 6.13.: *The RMSD trajectory at folding equilibrium temperature features many transitions between the folded and the unfolded state. Roughly every* 1 *million MTM-G steps the system performs a transition, which allows us to precisely determine barriers between free energy states. The native conformation could be found multiple times with an error below* 1Å *backbone RMSD to the crystal structure, excluding the unstructured residues* $32 - 36$*.*

6.3.3. Multiple Try Metropolis

After showing in chapter 5 that the MTM-G method can indeed be used for modeling protein systems, we apply this parallel approach to the investigation of the ultrafast folding Villin headpiece subdomain. Within a single simulation in the atomistic AMBER99SB-STAR-ILDN force field comprising $6.3 \cdot 10^7$ steps we could observe 135 transitions between the unfolded state with $Q <= 0.2$ and the folded state located at $Q >= 0.75$ (Figure 6.13). With this incredible transition frequency (in comparison the MH algorithm required 10^8 steps for 4 transitions, see section 6.3.1), the free energy landscape of the system can be determined with high accuracy. The accuracy for refolding to the native configuration, determined by calculating the backbone RMSD excluding the unstructured residues $32 - 36$ with respect to the experimentally resolved crystal structure, is very high and exhibits an error below to 1Å. The free energy profile extracted from this simulation (see section 6.3.1 for details), perfectly fits to the accumulated free energy landscape of all MH simulations (Figure 6.14) and thus the simulation produced the same thermodynamic expectation values. Besides the folded and the unfolded state the MTM-G simulation also revealed the folding intermediate state and the stabilization of the folded state by a free energy barrier of approximately $1\frac{kcal}{mol}$. The tree distinct states can also be observed in an energy vs. reaction coordinate plot (Figure 6.15). The phase space of the protein system was exhaustively sampled and the RMSD as well as the Q are well correlated to the system's potential energy. In both reaction coordinates, the minimal potential energy state represents well the folded state at 1Å RMSD and 0.82 Q. Using an additional reaction coordinate R_g, the radius of gyration, we can calculate a two-dimensional potential of mean force (PMF) (Figure 6.16). All three states are represented by pronounced peaks and connected with each other by lower populated transition pathways. By comparing to a experimental folding times of $8\mu s$, a single

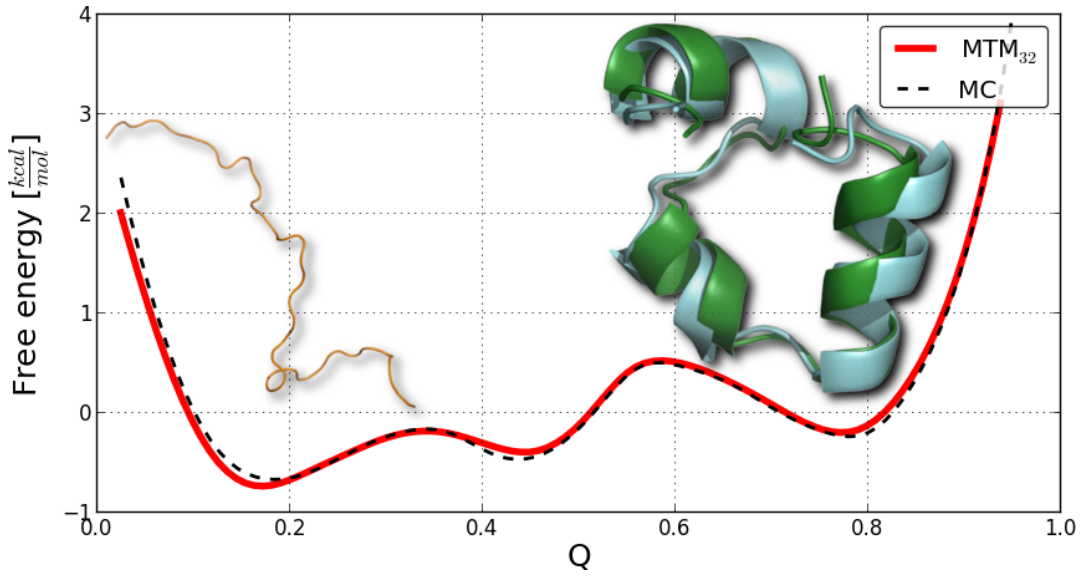


Figure 6.14.: *The resulting free energy profile, calculated from a single MTM trajectory equals the aggregated free energy profile of the previous MH simulations. Both profiles exhibit a three state system including the unfolded ($Q = 0.17$), an intermediate ($Q = 0.46$) and the folded state ($Q = 0.78$), which is stabilized by a barrier of $1\frac{kcal}{mol}$. One conformation from the unfolded ensemble is shown left (orange) and the overlay of the folded conformation (blue) with the crystal structure (green) is shown right.*

MTM-G$_{32}$ step translates to a time equivalent of approximately $17ps$. Current simulations can accomplish roughly 10^6 steps per day on a 128 CPU machine and thus produce results, comparable to those of the Anton supercomputer, which lie in the order of $20\mu s$ per day. However, the use of our MTM-G method requires 17 times less computational effort (Section 5.3.3) and thus facilitates the general investigation of complex processes on long time scales. With roughly two folding / unfolding transitions per day simulation time, the MTM-G allows the description of the protein folding process within a single day on conventional computer architectures.

6.4. Discussion

We could demonstrate that the thermodynamic description of proteins using a Monte Carlo approach allows the complete thermodynamic characterization of protein folding, while consuming much less computing time than an MD approach. Using SIMONA we could successfully characterize the ultrafast folding subdomain of the Villin headpiece protein and determine its thermodynamic properties by performing multiple MH simulations on off the shelf hardware. The observation of multiple folding and unfolding events allowed us to determine the free energy profile of this protein and hence revealed a folding intermediate state. Comparing the frequency of these transitions with experimental folding times translates to a high simulation efficiency with a time equivalent of $116fs$ per MH step. We could measure a stability of $1\frac{kcal}{mol}$ for the native conformation. The calculated CD spectra reproduce experimental observations very well and grants insight into the secondary structure of the simulated protein at a given temperature. Further we could determine the temperature dependent ellipticity of the system, which would require enormous computational effort using molecular dynamics approaches and was never done before. With our MH simulations, we are capable of giving information about

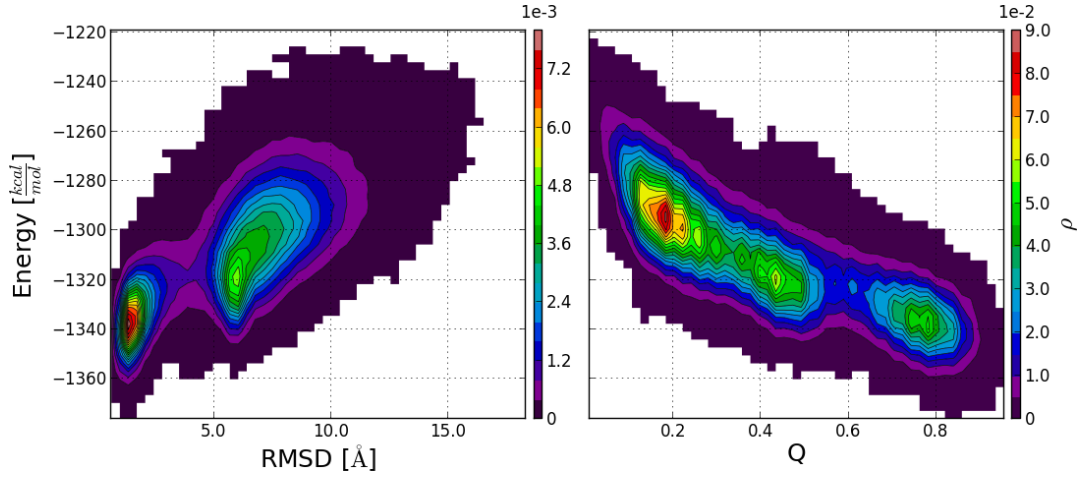


Figure 6.15.: *Both histograms show distinct peaks for the folded and the unfolded state. However, only the more accurate Q allows the observation of a folding intermediate in between of them. The minimal potential energy lies, with* 1Å *RMSD /* 0.82 *Q, well at the folded state.*

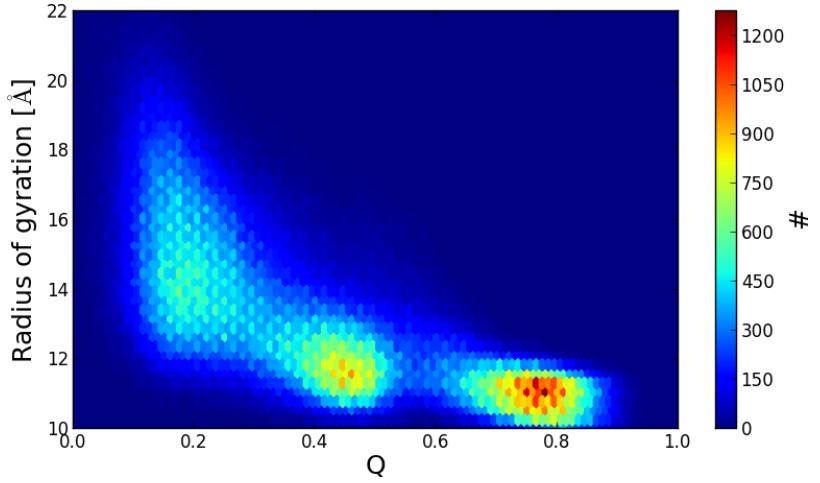


Figure 6.16.: *In the two-dimensional potential of mean force (PMF) spanned by Q and a second reaction coordinate, the radius of gyration, all three states are visible in distinct peaks. Transitions occur along lower populated pathways, which form connections between the higher occupied minima.*

the structural properties of single amino acids within the transition state to either the native or the unfolded ensemble.

By allowing structural interchange between different temperatures in a parallel tempering simulation scheme, we could characterize thermodynamic observables of the same protein within a simulation of 32 temperature basins. Using properly chosen temperature ranges, the simulations could overcome the problem of being frozen within local minima and thus increase the sampling efficiency.

Using the massively parallel Monte Carlo simulation method MTM-G developed in this work, the subdomain of the Villin headpiece could be exhaustively sampled and folding transition can be observed within a single day of simulation. The calculated time equivalent based on the observed transition counts and the experimental folding time of the protein suggests an efficiency of $17ps$ per MTM-G step, which would equal $\approx 10^4$ MD steps. This incredible acceleration offers exciting new possibilities to investigate most complex biological processes in atomistic simulations.

In further investigations, the single temperature MTM-G simulations could be mixed with a multi-temperature PT approach to perform hybrid PT-MTM simulations. We think that those simulation could grant even more enhancements for the characterization of protein systems.

7. Summary

Proteins constitute the major part of a cell's dry matter and perform many essential functions in all living organisms. In addition to their structural contribution they accomplish a multitude of cell and body functions. In an evolutionary process, their unique structure has evolved to perfectly fit their associated functionalities. Even small structural changes, caused for example by mutations or misfolding, may lead to diseases. For the development of proper therapies, the functional behavior of the proteins involved needs to be understood and the knowledge of the structure of a protein is often the key to its functionality. However, the experimental determination of the protein structures often proves difficult. As a result, compared to the vast number of known sequences, only relatively few structures have been resolved so far. Especially for membrane proteins, which constitute 50% of current drug targets, the protein database only comprises 237 high resolution structures. Embedding in the cell membrane complicates methods to obtain structures of membrane proteins with atomic resolution. Characterization in artificial environments is equally challenging due to difficulties in expression, damage during extraction, refolding of the protein, or failure to crystallize. In addition, structural changes induced by the artificial environment may lead to the observation of non-native conformations.

In addition to the importance of static structures, it is known that protein dynamics are crucial for function. Application of structure determination methods like X-ray diffraction is extremely challenging to give insights into these protein dynamics and computer based methods have long been investigated to elucidate the connection between the protein flexibility and function. The predominant simulation technique, Molecular Dynamics (MD), exhibits the time scale problem, which severely restricts the applicability of the simulation method and prevents the investigation of large scale conformational changes, like protein folding, assembly or the interaction of larger complexes. The advances in computer design have recently enabled the characterization of the structural dynamics of ultrafast folding proteins, but these simulations still remain very expensive and are limited to a single group worldwide.

In this thesis, I have investigated a Monte Carlo based approach to avoid the simulation of short time scale atomic vibrations, which nevertheless enables the simulation of the thermodynamic properties of the systems. In addition, I have developed a coarse-grained, implicit model, which incorporates experimental observations, to investigate a novel mechanism as the driving force behind the assembly of membrane proteins.
After implementing them in an efficient simulation package, I have been able to demonstrate that Monte Carlo (MC) based simulation techniques can efficiently describe thermodynamic properties of protein systems in excellent agreement with experimental observations and prior all-atom explicit solvent molecular dynamics simulations. Using this stochastic strategy, we directly constructed a representative conformational ensemble without having explored irrelevant fast transitions like atomic vibrations. We have also succeeded to develop massively parallel

MC based sampling techniques to make the application of modern highly parallel computer architectures to biomolecular simulations possible. Our modified Multiple Try Metropolis (MTM) method allows the investigation of reversible folding and unfolding transitions of ultrafast folding proteins within a single day.

We developed SIMONA[2, 3]: *an efficient and versatile framework for stochastic simulations of molecular and nanoscale systems*, which has been lacking among the available simulation methods, which strongly focus on molecular dynamics simulations. The simulation package is freely available for academic use and can be downloaded at: `http://int.kit.edu/nanosim/simona`.

Twin Arginine Translocase

Translocation of proteins across cell membranes is a fundamental mechanism of cells. A particular interesting transport mechanism is the Twin Arginine Translocase (Tat), because it transports completely folded proteins. Unlike conventional transport systems, which transport unstructured chains of amino acids, the Tat system forms much larger pores selective to each cargo protein. Especially in bacteria, which rely on photosynthetic electron transfer, the transport of cofactor binding proteins is essential. But binding a cofactor requires the protein to fold in the cytosol and thus the protein cannot be translocated by other transport mechanisms, such as the Sec pathway. The Tat mechanism comprises two different protein types TatA and TatC in various quantities to accommodate the size of the cargo protein. Depending on the organism, further TatA homologs, such as the TatB homolog, may contribute to the translocation process. The heterogeneity of assembled TatA pores complicates an experimental investigation and only low resolution data is available to date.

In collaboration with the group Ulrich at the KIT and the group Ruggerone at the Università di Cagliari we investigated the sequential information of the $\mathrm{TatA_d}$ protein and observed an intrinsic complementary charge pattern along adjacent structural parts, which would allow a zipper like formation of salt bridges[4]. Despite a total charge of nearly zero, TatA monomers comprise 40% charged residues in extra-membranous domains, 17.5% more than expected for transport proteins. The charge zipper motif within the $\mathrm{TatA_d/C_d}$ mechanism in the bacillus subtilis organism consists of seven pairwise complementary salt bridges between two different structural domains.
We postulated a successive formation of the intramolecular zipper contacts on a basis of structural models. Our model demonstrates the monomers to adopt a hairpin like structure, which would then be able to aggregate via intermolecular salt bridges to amphiphilic palisades with adequate heights to span the membrane. In a trapdoor like mechanism, such amphiphilic palisades could form translocation pores with a hydrophilic interior spanning the membrane. In addition to TatA, we found putative charge zipper motifs in 75% of all membrane proteins in the Uniprot database. This led to our assumptions that the charge zipper mechanism is a fundamental mechanism for the self-assembly of membrane proteins.

I developed a structural, implicit membrane-pore model including explicit attractive constraints

to model the proposed salt bridges and elucidated the importance of charge zipper motifs for the self-assembly process and the structural constitution of $TatA_d$ pores. In extensive mutation studies, our experimental partners confirmed the existence of the salt-bridge pattern consisting of five intramolecular and two intermolecular contacts. Based on this motif, we successfully simulated the self-assembly process of dodecameric $TatA_d$ pores, in very good agreement with experimental observations. The shape and diameter of the simulated pore systems agreed well with cryo-EM measurements. Our simulations provide the first atomic model of $TatA_d$ pores and give insights into the self-assembly process of the TatA mechanism.

By simulating the aggregation of the bacterial stress-response TisB peptide, the formation of tetrameric pore complexes of the anionic antimicrobial Dermcidin peptide and the membrane alignment of the structural glycoprotein of pestiviruses, E^{rns}, we were able to demonstrate the transferability of our approach.

Multiple Try Metropolis

To facilitate the simulation of long time scale processes in the TatA mechanism, we coarse grained atomic interactions and thereby eliminated the irrelevant fast transitions, which are responsible for the time scale problem of an MD simulation. Without coarse graining, the femtosecond integration step in combination with sophisticated force fields, restricts MD simulations to realms below the microsecond time scale. Nevertheless, simulations based on physical energy functions are required for accurate predictions of biological function without further knowledge of the systems. But most interesting processes occur on long time scales, which are unreachable using MD techniques. Stochastic Monte Carlo methods overcome the time scale problem by generating the stationary distribution of the system to measure thermodynamic expectation values without incorporating atomic vibrations. However, MC and MD simulations are strictly sequential techniques, which cannot be executed in parallel and accelerations only rely on parallelization of the energy evaluation.
An extension to the common Metropolis Hastings MC method (MH), the Multiple Try Metropolis method (MTM) modifies the proposal step of MH, which would otherwise generate a single random conformation, to generate multiple trial conformations at once. This seems to cause a large computational overhead, but nevertheless, for mathematical problems, the more complex exploration of the phase space converges faster than conventional MH approaches. Even by taking the computational overhead into account, MTM attains smaller auto-correlation times than the standard MH and thus provides a more efficient traversal of the phase space. Yet, it was never successfully applied to the field of protein folding.
In this work I investigated the applicability of the MTM method to the complex potential functions of proteins. I successfully developed a modified MTM-G method, which, in contrast to the standard MTM, is indeed able to efficiently perform simulations of protein systems. The introduction of modified transition probabilities allows robust simulation of proteins consistent with the detailed balance criterion and thus produces correct thermodynamic expectation values. Protein trajectories, generated by an MTM-G method, exhibit much smaller auto-correlation times compared to conventional MH simulation. This promises a faster phase space exploration

speed and a more efficient characterization of thermodynamic properties. In a closer examination of the MTM-G method I demonstrated its parallelization capabilities at a high simulation efficiency. Compared to sequential simulation techniques like MH or MD, the MTM-G allows the parallel simulation of a single trajectory. This permits the application on supercomputers and thereby facilitates the investigation of long time scale processes. Thus, the MTM-G method provides an excellent parallel extension to the MH method without reducing the complexity of the simulated systems.

Protein folding

For many proteins the three-dimensional structure is completely determined by the primary amino acid sequence as most proteins fold spontaneously, or with the help of other proteins, from an unstructured coil into the native conformation. In most cases, the folding process is reversible and, depending on the environmental conditions, takes place on microseconds to second time scales. For many problems in life-sciences the knowledge of the 3D-structure of a protein is very important, as it is the most versatile basis for the elucidation of biological processes. In addition to experimental investigations, theoretic methods increasingly complement experiments to the understanding of protein function. The most widely used simulation technique, molecular dynamics, is strongly limited by the time scale problem and thus not suitable to explore processes occurring on long timescales. Stochastic MC approaches are a convenient alternative, but not many modern MC simulation programs exist for the biological field.
As part of this work, we have developed the efficient Monte Carlo based simulation framework SIMONA, in which all Metropolis based techniques discussed in this thesis were implemented. Using a conventional MH simulation approach we successfully performed reversible protein folding simulations of the subdomain of the Villin headpiece protein. We were able to completely characterize its thermodynamic properties with atomic resolution. The calculation of its free energy landscape revealed a folding intermediate, separated from the native configuration by a barrier of $1\frac{kcal}{mol}$. We could provide information about the thermostability of single amino acids as well as helical segments. In the early structure-formation process, we observed the formation of helical tertiary contacts consistent with the nucleation model. By analyzing the per-residue ϕ-values, we could characterize the folding-transition state, where the third helix is mostly unfolded. We could characterize the secondary structure by calculating the averaged circular dichroism spectra over all conformations. This allowed us to determine the temperature dependent ellipticity in good agreement with experimental observations. For this complete investigation we required around 180 times less computational effort on off-the-shelf hardware compared to a $100\mu s$ MD simulation on the Anton supercomputer.
By allowing structural interchanges within multiple temperature basins in a replica exchange simulation, we were able to perform a parallel characterization of the folding landscape of the same protein using less computational effort. The exchange between distinct temperatures increases the barrier crossing rate for every single temperature and thus results in a more accurate description of the underlying free energy profile.
With the application of the MTM-G method developed in this thesis, we further enhance the simulation efficiency obtaining far more folding and unfolding transitions compared to con-

ventional MH simulations. Using massively parallel computer architectures, we were able to perform reversible folding simulations of the Villin headpiece subdomain within a single day. The resulting free energy landscapes perfectly matched those from previous MH simulations. However, the generated conformational ensembles exhibit much lower auto-correlation times and thus provide more efficient determination of thermodynamic expectation values. Comparing the folding frequency with experimentally resolved folding times translates to an efficiency of $17ps$ per MTM-G step, which lies more than four orders of magnitude above an MD step. This immensely increased simulation speed offers many exciting opportunities to investigate biological processes beyond the time scales reachable by MD approaches.

A. Appendix

A.1. SIMONA - Source Code Statistics

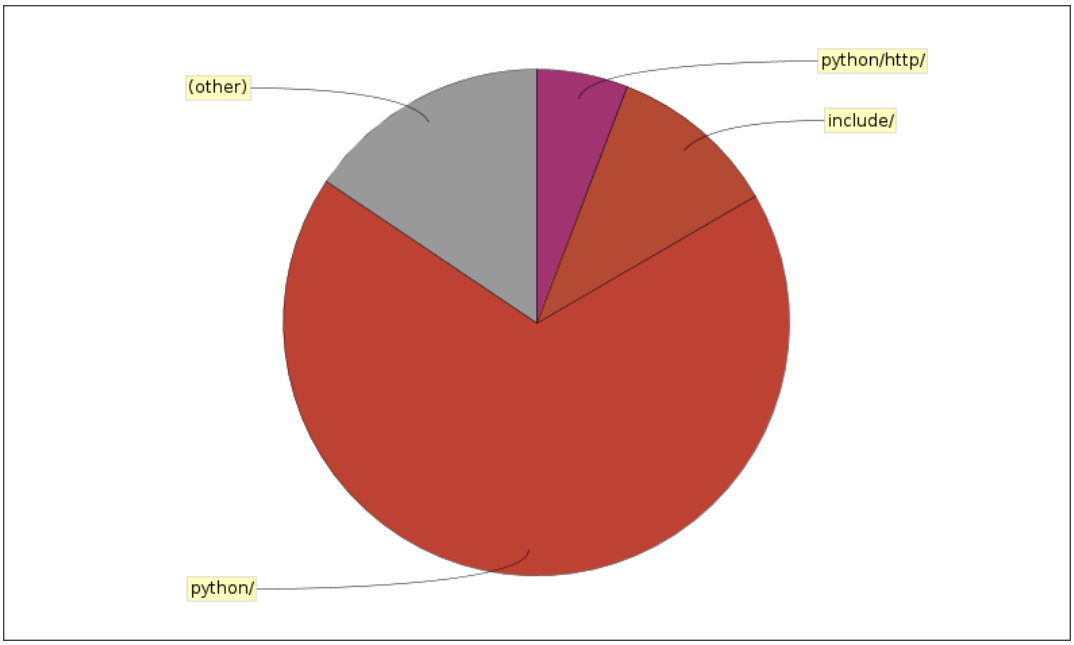


Figure A.1.: *My contribution was focused on the preprocessor (python), the multi transformations (other, include) and web-based applications for SIMONA.*

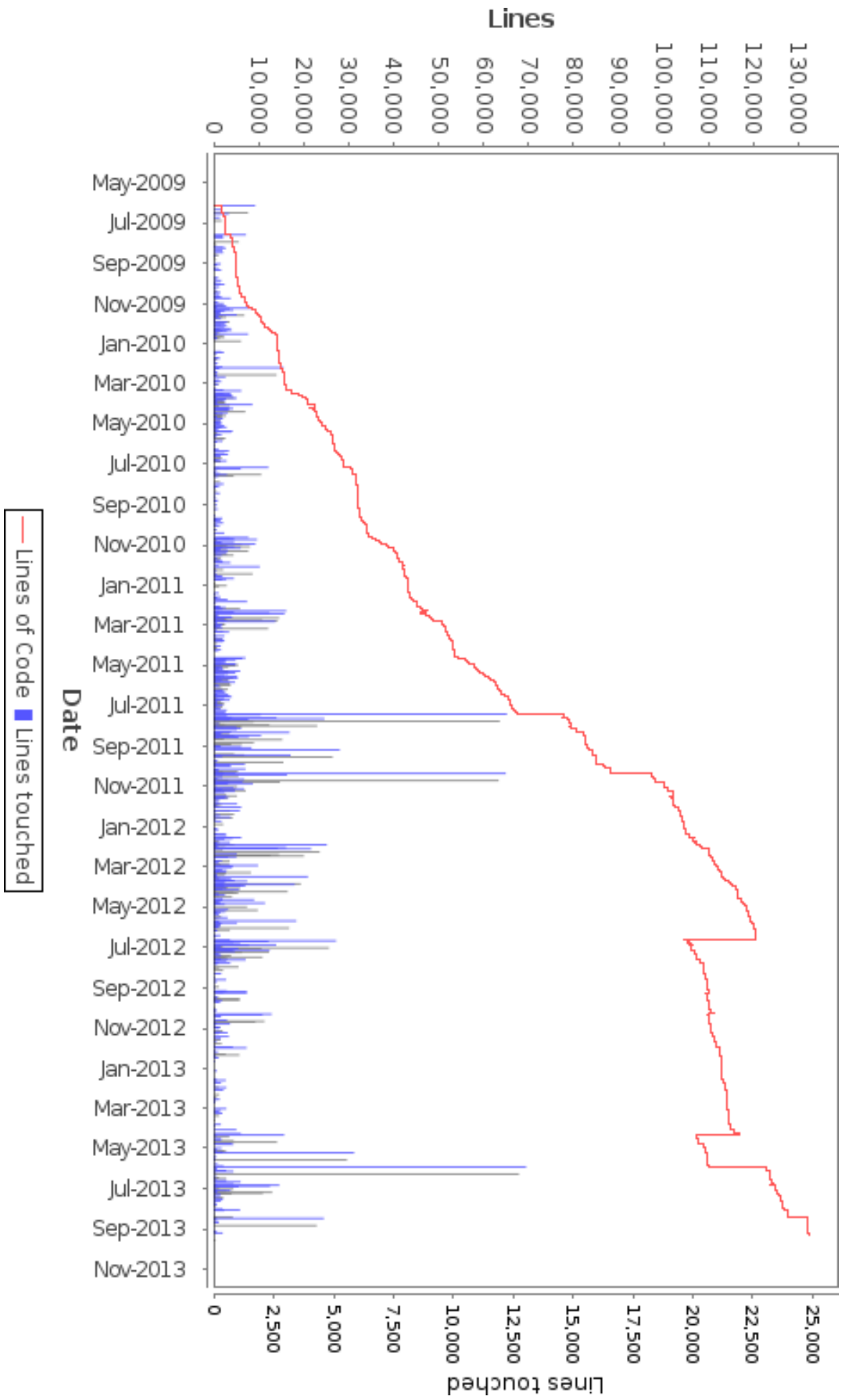


Figure A.2.: *Time evolution of the number of lines of code (LoC) included in the SIMONA package.*

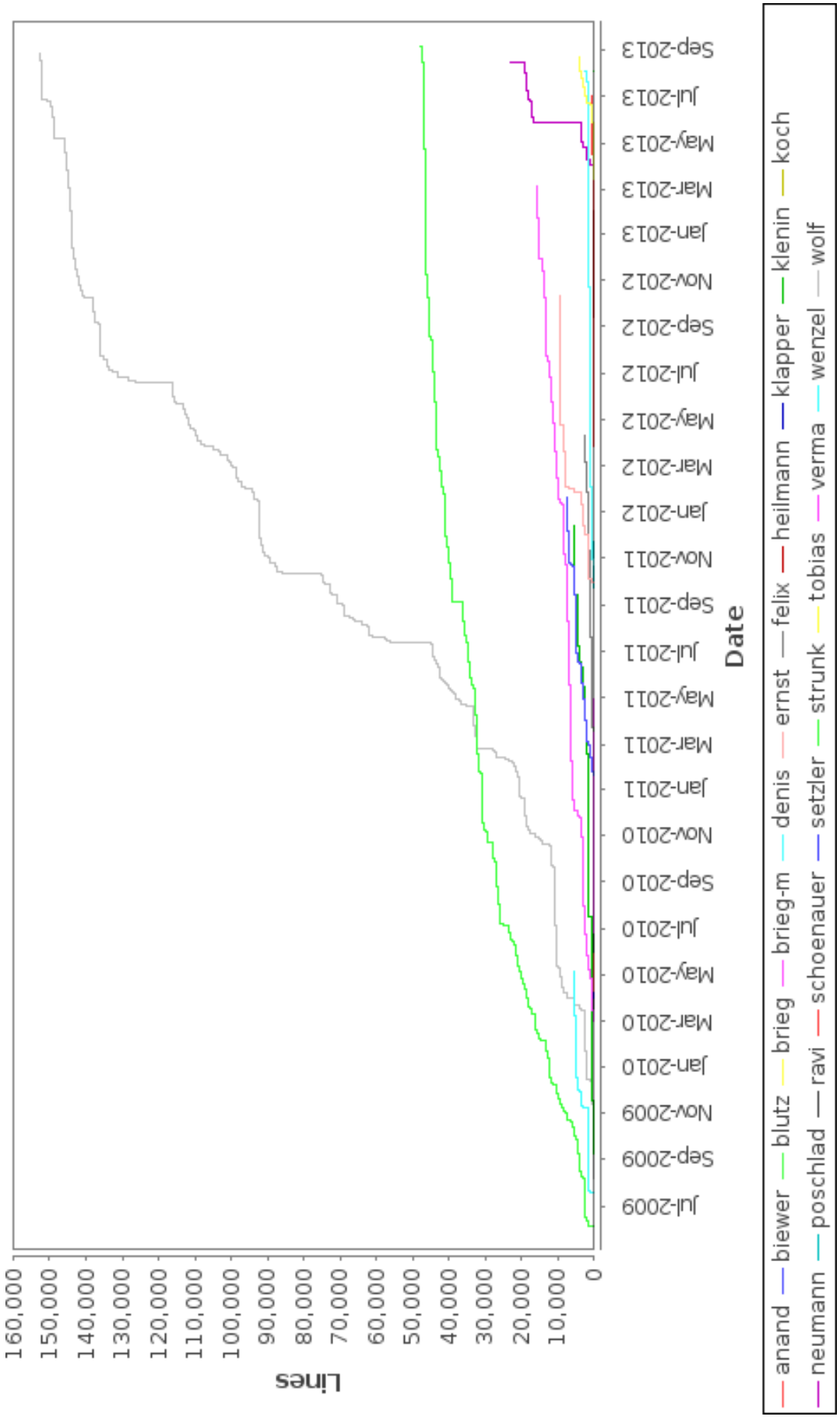
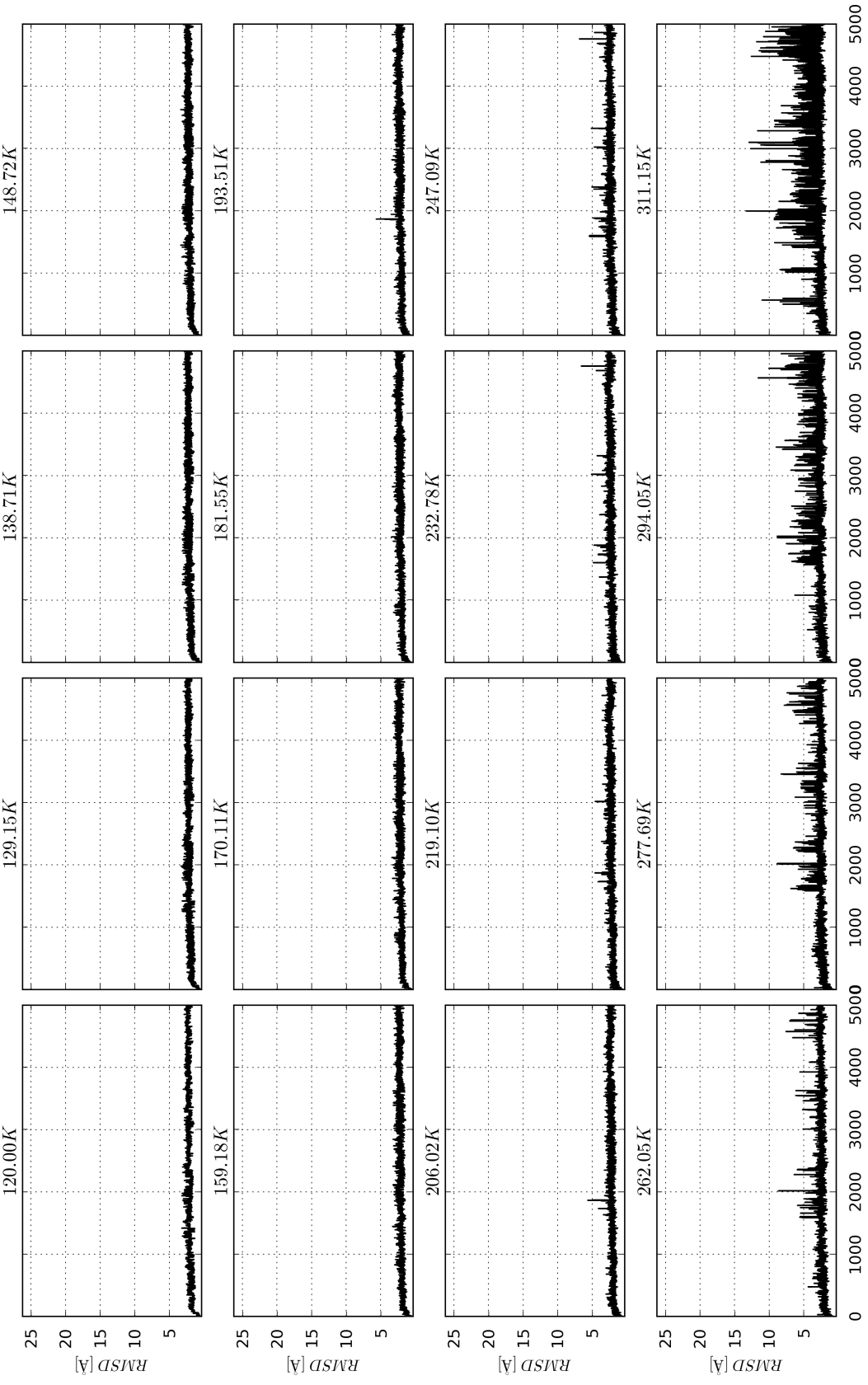
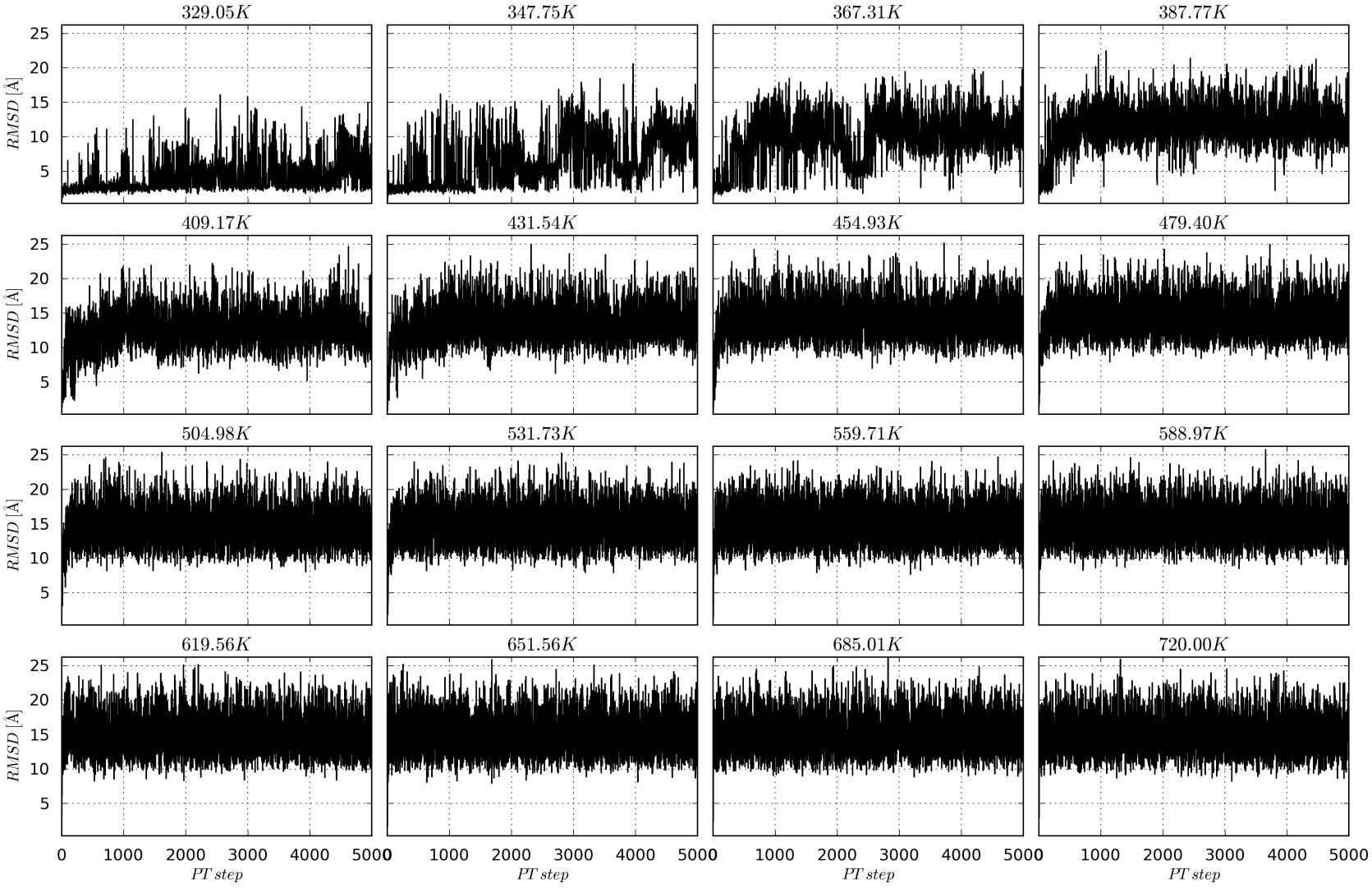


Figure A.3.: *Source code contribution per author.*

A.2. PT - RMSD plots



(a) *PT: RMSD plot - part 1.*



(b) *PT: RMSD plot - part 2.*

Figure A.2.: *RMSD vs. PT-step plots for each one of the* 32 *exponentially distributed temperatures. The folding equilibrium temperatures lie at* $\approx 354K$*. The system requires a first* 2000 *PT steps to equilibrate the initial structure to the simulation temperature.*

List of Figures

Bibliography

[1] Hiroyuki Mori and Koreaki Ito. "The Sec protein-translocation pathway". In: *Trends in Microbiology* 9.10 (Oct. 2001), pp. 494–500. ISSN: 0966-842X. DOI: 10.1016/S0966-842X(01)02174-6. URL: http://www.cell.com/trends/microbiology/abstract/S0966-842X(01)02174-6.

[2] M Wolf et al. "SIMONA 1.0: an efficient and versatile framework for stochastic simulations of molecular and nanoscale systems". eng. In: *Journal of computational chemistry* 33.32 (Dec. 2012), pp. 2602–2613. ISSN: 1096-987X. DOI: 10.1002/jcc.23089.

[3] T. Strunk, M. Wolf, and W. Wenzel. "Development and evaluation of a GPU-optimized N-body term for the simulation of biomolecules". In: *Computational Methods in Science and Engineering $ dProceedings of the Workshop SimLabs@ KIT, November 29-30, 2010, Karlsruhe, Germany*. 2010, p. 35.

[4] Torsten H. Walther et al. "Folding and Self-Assembly of the TatA Translocation Pore Based on a Charge Zipper Mechanism". In: *Cell* 152.1 (Jan. 2013), pp. 316–326. ISSN: 0092-8674. DOI: 10.1016/j.cell.2012.12.017. URL: http://www.cell.com/abstract/S0092-8674(12)01539-5.

[5] Bruce Alberts et al. *Molecular Biology of the Cell*. en. Text. 2002. URL: http://www.ncbi.nlm.nih.gov/books/NBK21054/.

[6] M. F. Perutz. "Hemoglobin Structure and Respiratory Transport". In: *Scientific American* 239.6 (Dec. 1978), pp. 92–125. ISSN: 0036-8733. DOI: 10.1038/scientificamerican1278-92. URL: http://adsabs.harvard.edu/abs/1978SciAm.239f..92P.

[7] L PAULING and R B COREY. "Atomic coordinates and structure factors for two helical configurations of polypeptide chains". In: *Proceedings of the National Academy of Sciences of the United States of America* 37.5 (May 1951), pp. 235–240. ISSN: 0027-8424. URL: http://www.ncbi.nlm.nih.gov/pubmed/14834145.

[8] L PAULING and R B COREY. "The pleated sheet, a new layer configuration of polypeptide chains". In: *Proceedings of the National Academy of Sciences of the United States of America* 37.5 (May 1951), pp. 251–256. ISSN: 0027-8424. URL: http://www.ncbi.nlm.nih.gov/pubmed/14834147.

[9] L PAULING and R B COREY. "The structure of feather rachis keratin". In: *Proceedings of the National Academy of Sciences of the United States of America* 37.5 (May 1951), pp. 256–261. ISSN: 0027-8424. URL: http://www.ncbi.nlm.nih.gov/pubmed/14834148.

[10] L PAULING and R B COREY. "The structure of fibrous proteins of the collagen-gelatin group". In: *Proceedings of the National Academy of Sciences of the United States of America* 37.5 (May 1951), pp. 272–281. ISSN: 0027-8424. URL: http://www.ncbi.nlm.nih.gov/pubmed/14834150.

[11] L PAULING and R B COREY. "The structure of hair, muscle, and related proteins". In: *Proceedings of the National Academy of Sciences of the United States of America* 37.5 (May 1951), pp. 261–271. ISSN: 0027-8424. URL: http://www.ncbi.nlm.nih.gov/pubmed/14834149.

[12] L PAULING and R B COREY. “The structure of synthetic polypeptides”. In: *Proceedings of the National Academy of Sciences of the United States of America* 37.5 (May 1951), pp. 241–250. ISSN: 0027-8424. URL: http://www.ncbi.nlm.nih.gov/pubmed/14834146.

[13] L PAULING, R B COREY, and H R BRANSON. “The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain”. In: *Proceedings of the National Academy of Sciences of the United States of America* 37.4 (Apr. 1951), pp. 205–211. ISSN: 0027-8424. URL: http://www.ncbi.nlm.nih.gov/pubmed/14816373.

[14] C. Ramakrishnan and G. N. Ramachandran. “Stereochemical Criteria for Polypeptide and Protein Chain Conformations”. In: *Biophysical Journal* 5.6 (Nov. 1965). ISSN: 0006-3495.

[15] Jakob T Nielsen et al. “Unique identification of supramolecular structures in amyloid fibrils by solid-state NMR spectroscopy”. eng. In: *Angewandte Chemie (International ed. in English)* 48.12 (2009), pp. 2118–2121. ISSN: 1521-3773. DOI: 10.1002/anie.200804198.

[16] Jae Hyun Cho et al. “Insight into the stereochemistry in the inhibition of carboxypeptidase A with N-(hydroxyaminocarbonyl)phenylalanine: binding modes of an enantiomeric pair of the inhibitor to carboxypeptidase A”. eng. In: *Bioorganic & medicinal chemistry* 10.6 (June 2002), pp. 2015–2022. ISSN: 0968-0896.

[17] Chen Song et al. “Crystal structure and functional mechanism of a human antimicrobial membrane channel”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 110.12 (Mar. 2013), pp. 4586–4591. ISSN: 1091-6490. DOI: 10.1073/pnas.1214739110.

[18] C B Anfinsen. “Principles that govern the folding of protein chains”. In: *Science (New York, N.Y.)* 181.96 (July 1973), pp. 223–230. ISSN: 0036-8075. DOI: 10.1126/science.181.4096.223. URL: http://www.ncbi.nlm.nih.gov/pubmed/4124164.

[19] Cyrus Levinthal. “Are There Pathways For Protein Folding?” In: *Extrait du Journal de Chimie Physique* 65.1 (1968).

[20] Ken A. Dill and Hue Sun Chan. “From Levinthal to pathways to funnels”. In: *Nat Struct Mol Biol* 4.1 (Jan. 1997), pp. 10–19. DOI: 10.1038/nsb0197-10. URL: http://dx.doi.org/10.1038/nsb0197-10.

[21] J N Onuchic et al. “Toward an outline of the topography of a realistic protein-folding funnel”. In: *Proceedings of the National Academy of Sciences of the United States of America* 92.8 (Apr. 1995), pp. 3626–3630. ISSN: 0027-8424. URL: http://www.ncbi.nlm.nih.gov/pubmed/7724609.

[22] Timo Strunk et al. “Benchmarking the POEM@ HOME Network for Protein Structure Prediction”. In: *Proceedings of the 3rd International Workshop on Science Gateways for Life Sciences* (June 2011).

[23] Timo Strunk, Moritz Wolf, and Wolfgang Wenzel. “Peptide structure prediction using distributed volunteer computing networks”. en. In: *Journal of Mathematical Chemistry* 50.2 (Feb. 2012), pp. 421–428. ISSN: 0259-9791, 1572-8897. DOI: 10.1007/s10910-011-9937-x. URL: http://link.springer.com/article/10.1007/s10910-011-9937-x.

[24] P S Kim and R L Baldwin. “Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding”. eng. In: *Annual review of biochemistry* 51 (1982), pp. 459–489. ISSN: 0066-4154. DOI: 10.1146/annurev.bi.51.070182.002331.

[25] O B Ptitsyn. “Stages in the mechanism of self-organization of protein molecules”. rus. In: *Doklady Akademii nauk SSSR* 210.5 (June 1973), pp. 1213–1215. ISSN: 0002-3264.

[26] Jayant B. Udgaonkar and Robert L. Baldwin. “NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A”. en. In: *Nature* 335.6192 (Oct. 1988), pp. 694–699. DOI: 10.1038/335694a0. URL: http://www.nature.com/nature/journal/v335/n6192/abs/335694a0.html.

[27] K A Dill. “Theory for the folding and stability of globular proteins”. eng. In: *Biochemistry* 24.6 (Mar. 1985), pp. 1501–1509. ISSN: 0006-2960.

[28] V R Agashe, M C Shastry, and J B Udgaonkar. “Initial hydrophobic collapse in the folding of barstar”. eng. In: *Nature* 377.6551 (Oct. 1995), pp. 754–757. ISSN: 0028-0836. DOI: 10.1038/377754a0.

[29] A. M. Gutin, V. I. Abkevich, and E. I. Shakhnovich. “Is Burst Hydrophobic Collapse Necessary for Protein Folding?” In: *Biochemistry* 34.9 (Mar. 1995), pp. 3066–3076. ISSN: 0006-2960. DOI: 10.1021/bi00009a038. URL: http://dx.doi.org/10.1021/bi00009a038.

[30] Donald B. Wetlaufer. “Nucleation, Rapid Folding, and Globular Intrachain Regions in Proteins”. en. In: *Proceedings of the National Academy of Sciences* 70.3 (Mar. 1973), pp. 697–701. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.70.3.697. URL: http://www.pnas.org/content/70/3/697.

[31] W Bolton and M F Perutz. “Three dimensional fourier synthesis of horse deoxyhaemoglobin at 2.8 Angstrom units resolution”. eng. In: *Nature* 228.5271 (Nov. 1970), pp. 551–552. ISSN: 0028-0836.

[32] B Shaanan. “Structure of human oxyhaemoglobin at 2.1 A resolution”. eng. In: *Journal of molecular biology* 171.1 (Nov. 1983), pp. 31–59. ISSN: 0022-2836.

[33] G Fermi et al. “The crystal structure of human deoxyhaemoglobin at 1.74 A resolution”. eng. In: *Journal of molecular biology* 175.2 (May 1984), pp. 159–174. ISSN: 0022-2836.

[34] Harvey Lodish et al. *Molecular Cell Biology*. en. Text. 2000. URL: http://www.ncbi.nlm.nih.gov/books/NBK21475/.

[35] Johann Deisenhofer and Hartmut Michel. “Structures of Bacterial Photosynthetic Reaction Centers”. In: *Annual Review of Cell Biology* 7.1 (1991), pp. 1–23. DOI: 10.1146/annurev.cb.07.110191.000245. URL: http://www.annualreviews.org/doi/abs/10.1146/annurev.cb.07.110191.000245.

[36] R. Henderson et al. “Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy”. In: *Journal of Molecular Biology* 213.4 (June 1990), pp. 899–929. ISSN: 0022-2836. DOI: 10.1016/S0022-2836(05)80271-2. URL: http://www.sciencedirect.com/science/article/pii/S0022283605802712.

[37] Hartmut Luecke et al. “Structure of bacteriorhodopsin at 1.55 Å resolution”. In: *Journal of Molecular Biology* 291.4 (Aug. 1999), pp. 899–911. ISSN: 0022-2836. DOI: 10.1006/jmbi.1999.3027. URL: http://www.sciencedirect.com/science/article/pii/S0022283699930279.

[38] B. J. Alder and T. E. Wainwright. “Phase Transition for a Hard Sphere System”. In: *The Journal of Chemical Physics* 27.5 (Nov. 1957), pp. 1208–1209. ISSN: 00219606. DOI: doi:10.1063/1.1743957. URL: http://jcp.aip.org/resource/1/jcpsa6/v27/i5/p1208_s1?isAuthorized=no.

[39] B. J. Alder and T. E. Wainwright. “Studies in Molecular Dynamics. I. General Method”. In: *The Journal of Chemical Physics* 31.2 (Aug. 1959), pp. 459–466. ISSN: 00219606. DOI: doi:10.1063/1.1730376. URL: http://jcp.aip.org/resource/1/jcpsa6/v31/i2/p459_s1.

[40] A. Rahman. “Correlations in the Motion of Atoms in Liquid Argon”. In: *Physical Review* 136.2A (Oct. 1964), A405–A411. DOI: 10.1103/PhysRev.136.A405. URL: http://link.aps.org/doi/10.1103/PhysRev.136.A405.

[41] Aneesur Rahman and Frank H. Stillinger. “Molecular Dynamics Study of Liquid Water”. In: *The Journal of Chemical Physics* 55.7 (Oct. 1971), pp. 3336–3359. ISSN: 00219606. DOI: doi:10.1063/1.1676585. URL: http://jcp.aip.org/resource/1/jcpsa6/v55/i7/p3336_s1?isAuthorized=no.

[42] Frank H. Stillinger and Aneesur Rahman. “Improved simulation of liquid water by molecular dynamics”. In: *The Journal of Chemical Physics* 60.4 (Feb. 1974), pp. 1545–1557. ISSN: 00219606. DOI: doi:10.1063/1.1681229. URL: http://jcp.aip.org/resource/1/jcpsa6/v60/i4/p1545_s1?isAuthorized=no.

[43] Kai Kadau, Timothy C. Germann, and Peter S. Lomdahl. “Large-Scale Molecular-Dynamics Simulation of 19 Billion Particles”. In: *International Journal of Modern Physics C* 15.01 (Jan. 2004), pp. 193–201. ISSN: 0129-1831, 1793-6586. DOI: 10.1142/S0129183104005590. URL: http://www.worldscientific.com/doi/abs/10.1142/S0129183104005590.

[44] Peter S. Lomdahl, Timothy C. Germann, and Kai Kadau. “Multibillion-atom Molecular Dynamics Simulations on BlueGene/L”. In: *Bulletin of the American Physical Society*. American Physical Society, Aug. 2005. URL: http://meetings.aps.org/link/BAPS.2005.SHOCK.U5.3.

[45] Tetsu Narumi et al. “A 55 TFLOPS simulation of amyloid-forming peptides from yeast prion Sup35 with the special-purpose computer system MDGRAPE-3”. In: *Proceedings of the 2006 ACM/IEEE conference on Supercomputing*. SC ’06. New York, NY, USA: ACM, 2006. ISBN: 0-7695-2700-0. DOI: 10.1145/1188455.1188506. URL: http://doi.acm.org/10.1145/1188455.1188506.

[46] Peter L Freddolino et al. “Molecular dynamics simulations of the complete satellite tobacco mosaic virus”. eng. In: *Structure (London, England: 1993)* 14.3 (Mar. 2006), pp. 437–449. ISSN: 0969-2126. DOI: 10.1016/j.str.2005.11.014.

[47] Paul C Whitford et al. “Accommodation of aminoacyl-tRNA into the ribosome involves reversible excursions along multiple pathways”. eng. In: *RNA (New York, N.Y.)* 16.6 (June 2010), pp. 1196–1204. ISSN: 1469-9001. DOI: 10.1261/rna.2035410.

[48] Chao Mei et al. “Enabling and scaling biomolecular simulations of 100 million atoms on petascale machines with a multicore-optimized message-driven runtime”. In: *High Performance Computing, Networking, Storage and Analysis (SC), 2011 International Conference for*. 2011, pp. 1–11.

[49] Loup Verlet. “Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules”. In: *Physical Review* 159.1 (July 1967), pp. 98–103. DOI: 10.1103/PhysRev.159.98. URL: http://link.aps.org/doi/10.1103/PhysRev.159.98.

[50] R. W. Hockney. “The potential calculation and some applications”. English. In: (Jan. 1970). URL: http://ntrs.nasa.gov/search.jsp?R=19700060674.

[51] H. J. C. Berendsen et al. “Molecular dynamics with coupling to an external bath”. In: *The Journal of Chemical Physics* 81.8 (Oct. 1984), pp. 3684–3690. ISSN: 00219606. DOI: 10.1063/1.448118. URL: http://jcp.aip.org/resource/1/jcpsa6/v81/i8/p3684_s1.

[52] A. A. Markov. “An Example of Statistical Investigation of the Text Eugene Onegin Concerning the Connection of Samples in Chains”. In: *Science in Context* 19.04 (2006), pp. 591–600. DOI: 10.1017/S0269889706001074.

[53] Olle Häggström. *Finite Markov chains and algorithmic applications*. Cambridge University Press, 2002. ISBN: 9780521890014.

[54] Nicholas Metropolis et al. “Equation of State Calculations by Fast Computing Machines”. In: *The Journal of Chemical Physics* 21.6 (1953), p. 1087. ISSN: 00219606. DOI: 10.1063/1.1699114. URL: http://link.aip.org/link/JCPSA6/v21/i6/p1087/s1&Agg=doi.

[55] A Markov. “Extension of the Limit Theorems of Probability Theory to a Sum of Variables Connected in a Chain”. In: *Dynamic Probabilistic Systems (Volume I: Markov Models)*. Ed. by R Howard. John Wiley & Sons, Inc., 1971, pp. 552–577.

[56] Silvia Pandolfi, Francesco Bartolucci, and Nial Friel. “A generalization of the Multiple-try Metropolis algorithm for Bayesian estimation and model selection”. In: *Journal of Machine Learning Research - Proceedings Track* (2010), pp. 581–588.

[57] Jun S. Liu. *Monte Carlo Strategies in Scientific Computing*. URL: http://www.springer.com/statistics/statistical+theory+and+methods/book/978-0-387-76369-9.

[58] Erik Lindahl and Mark SP Sansom. “Membrane proteins: molecular dynamics simulations”. In: *Current Opinion in Structural Biology* 18.4 (Aug. 2008), pp. 425–431. ISSN: 0959-440X. DOI: 10.1016/j.sbi.2008.02.003. URL: http://www.sciencedirect.com/science/article/pii/S0959440X08000304.

[59] Erik Strandberg and Anne S. Ulrich. “NMR methods for studying membrane-active antimicrobial peptides”. en. In: *Concepts in Magnetic Resonance Part A* 23A.2 (2004), pp. 89–120. ISSN: 1552-5023. DOI: 10.1002/cmr.a.20024. URL: http://onlinelibrary.wiley.com/doi/10.1002/cmr.a.20024/abstract.

[60] Phillip J. Stansfeld and Mark S.P. Sansom. “Molecular Simulation Approaches to Membrane Proteins”. In: *Structure* 19.11 (Nov. 2011), pp. 1562–1572. ISSN: 0969-2126. DOI: 10.1016/j.str.2011.10.002. URL: http://www.sciencedirect.com/science/article/pii/S0969212611003364.

[61] Marcos Sotomayor and Klaus Schulten. “Molecular Dynamics Study of Gating in the Mechanosensitive Channel of Small Conductance MscS”. In: *Biophysical Journal* 87.5 (Nov. 2004), pp. 3050–3065. ISSN: 0006-3495. DOI: 10.1529/biophysj.104.046045. URL: http://www.sciencedirect.com/science/article/pii/S0006349504737765.

[62] Andriy Anishkin and Sergei Sukharev. “Water Dynamics and Dewetting Transitions in the Small Mechanosensitive Channel MscS”. In: *Biophysical Journal* 86.5 (May 2004), pp. 2883–2895. ISSN: 0006-3495. DOI: 10.1016/S0006-3495(04)74340-4. URL: http://www.sciencedirect.com/science/article/pii/S0006349504743404.

[63] William L. Jorgensen et al. “Comparison of simple potential functions for simulating liquid water”. In: *The Journal of Chemical Physics* 79.2 (July 1983), pp. 926–935. ISSN: 00219606. DOI: doi:10.1063/1.445869. URL: http://jcp.aip.org/resource/1/jcpsa6/v79/i2/p926_s1.

[64] HJC Berendsen et al. “Interaction models for water in relation to protein hydration”. In: *Intermolecular Forces* (1981), pp. 331–342.

[65] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. “The missing term in effective pair potentials”. In: *The Journal of Physical Chemistry* 91.24 (Nov. 1987), pp. 6269–6271. ISSN: 0022-3654. DOI: 10.1021/j100308a038. URL: http://dx.doi.org/10.1021/j100308a038.

[66] Jakob P. Ulmschneider and Martin B. Ulmschneider. “United Atom Lipid Parameters for Combination with the Optimized Potentials for Liquid Simulations All-Atom Force Field”. In: *Journal of Chemical Theory and Computation* 5.7 (July 2009), pp. 1803–1813. ISSN: 1549-9618. DOI: 10.1021/ct900086b. URL: http://dx.doi.org/10.1021/ct900086b.

[67] Anthony G. Lee. “Structural biology: Highly charged meetings”. en. In: *Nature* 462.7272 (Nov. 2009), pp. 420–421. ISSN: 0028-0836. DOI: 10.1038/462420a. URL: http://www.nature.com/nature/journal/v462/n7272/full/462420a.html.

[68] Jakob P. Ulmschneider et al. “Peptide Partitioning and Folding into Lipid Bilayers”. In: *Journal of Chemical Theory and Computation* 5.9 (Sept. 2009), pp. 2202–2205. ISSN: 1549-9618. DOI: 10.1021/ct900256k. URL: http://dx.doi.org/10.1021/ct900256k.

[69] James Gumbart et al. “Molecular dynamics simulations of proteins in lipid bilayers”. In: *Current opinion in structural biology* 15.4 (Aug. 2005). ISSN: 0959-440X. DOI: 10.1016/j.sbi.2005.07.007. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2474857/.

[70] Myunggi Yi, Hugh Nymeyer, and Huan-Xiang Zhou. “Test of the Gouy-Chapman Theory for a Charged Lipid Membrane against Explicit-Solvent Molecular Dynamics Simulations”. In: *Physical Review Letters* 101.3 (July 2008), p. 038103. DOI: 10.1103/PhysRevLett.101.038103. URL: http://link.aps.org/doi/10.1103/PhysRevLett.101.038103.

[71] B Roux and T Simonson. “Implicit solvent models”. eng. In: *Biophysical chemistry* 78.1-2 (Apr. 1999), pp. 1–20. ISSN: 0301-4622.

[72] B Roux. “Influence of the membrane potential on the free energy of an intrinsic protein.” In: *Biophysical Journal* 73.6 (Dec. 1997). ISSN: 0006-3495. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1181204/.

[73] John David Jackson. *Classical Electrodynamics*. URL: http://archive.org/details/ClassicalElectrodynamics.

[74] Nathan A Baker. “Poisson-Boltzmann methods for biomolecular electrostatics”. eng. In: *Methods in enzymology*. Vol. 383. 2004, pp. 94–118.

[75] Michael Feig, Wonpil Im, and 3rd Brooks Charles L. “Implicit solvation based on generalized Born theory in different dielectric environments”. eng. In: *The Journal of chemical physics* 120.2 (Jan. 2004), pp. 903–911. ISSN: 0021-9606. DOI: 10.1063/1.1631258.

[76] W. Clark Still et al. “Semianalytical treatment of solvation for molecular mechanics and dynamics”. In: *Journal of the American Chemical Society* 112.16 (Aug. 1990), pp. 6127–6129. ISSN: 0002-7863. DOI: 10.1021/ja00172a038. URL: http://dx.doi.org/10.1021/ja00172a038.

[77] Martin Brieg and Wolfgang Wenzel. “PowerBorn: A Barnes-Hut Tree Implementation for Accurate and Efficient Born Radii Computation”. In: *Journal of Chemical Theory and Computation* 9.3 (Mar. 2013), pp. 1489–1498. ISSN: 1549-9618. DOI: 10.1021/ct300870s. URL: http://dx.doi.org/10.1021/ct300870s.

[78] Alexey Onufriev, David A Case, and Donald Bashford. “Effective Born radii in the generalized Born approximation: the importance of being perfect”. eng. In: *Journal of computational chemistry* 23.14 (Nov. 2002), pp. 1297–1304. ISSN: 0192-8651. DOI: 10.1002/jcc.10126.

[79] Michael S Lee et al. “New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations”. eng. In: *Journal of computational chemistry* 24.11 (Aug. 2003), pp. 1348–1356. ISSN: 0192-8651. DOI: 10.1002/jcc.10272.

[80] Michael Feig et al. “Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures”. eng. In: *Journal of computational chemistry* 25.2 (Jan. 2004), pp. 265–284. ISSN: 0192-8651. DOI: 10.1002/jcc.10378.

[81] Harry A. Stern and Scott E. Feller. “Calculation of the dielectric permittivity profile for a nonuniform system: Application to a lipid bilayer simulation”. In: *The Journal of Chemical Physics* 118.7 (Feb. 2003), pp. 3401–3412. ISSN: 00219606. DOI: doi:10.1063/1.1537244. URL: http://jcp.aip.org/resource/1/jcpsa6/v118/i7/p3401_s1.

[82] Velin Z. Spassov, Lisa Yan, and Sándor Szalma. “Introducing an Implicit Membrane in Generalized Born/Solvent Accessibility Continuum Solvent Models”. In: *The Journal of Physical Chemistry B* 106.34 (Aug. 2002), pp. 8726–8738. ISSN: 1520-6106. DOI: 10.1021/jp020674r. URL: http://dx.doi.org/10.1021/jp020674r.

[83] Wonpil Im, Michael Feig, and Charles L. Brooks III. “An Implicit Membrane Generalized Born Theory for the Study of Structure, Stability, and Interactions of Membrane Proteins”. In: *Biophysical Journal* 85.5 (Nov. 2003), pp. 2900–2918. ISSN: 0006-3495. DOI: 10.1016/S0006-3495(03)74712-2. URL: http://www.sciencedirect.com/science/article/pii/S0006349503747122.

[84] Seiichiro Tanizaki and Michael Feig. “A generalized Born formalism for heterogeneous dielectric environments: application to the implicit modeling of biological membranes”. eng. In: *The Journal of chemical physics* 122.12 (Mar. 2005), p. 124706. ISSN: 0021-9606. DOI: 10.1063/1.1865992.

[85] Afra Panahi and Michael Feig. “Dynamic Heterogeneous Dielectric Generalized Born (DHDGB): An Implicit Membrane Model with a Dynamically Varying Bilayer Thickness”. In: *Journal of Chemical Theory and Computation* 9.3 (Mar. 2013), pp. 1709–1719. ISSN: 1549-9618. DOI: 10.1021/ct300975k. URL: http://dx.doi.org/10.1021/ct300975k.

[86] Seiichiro Tanizaki and Michael Feig. “Molecular Dynamics Simulations of Large Integral Membrane Proteins with an Implicit Membrane Model”. In: *The Journal of Physical Chemistry B* 110.1 (Jan. 2006), pp. 548–556. ISSN: 1520-6106. DOI: 10.1021/jp054694f. URL: http://dx.doi.org/10.1021/jp054694f.

[87] Siewert J. Marrink, Alex H. de Vries, and Alan E. Mark. “Coarse Grained Model for Semiquantitative Lipid Simulations”. In: *The Journal of Physical Chemistry B* 108.2 (Jan. 2004), pp. 750–760. ISSN: 1520-6106. DOI: 10.1021/jp036508g. URL: http://dx.doi.org/10.1021/jp036508g.

[88] Steve O. Nielsen et al. “Coarse grain models and the computer simulation of soft materials”. en. In: *Journal of Physics: Condensed Matter* 16.15 (Apr. 2004), R481. ISSN: 0953-8984. DOI: 10.1088/0953-8984/16/15/R03. URL: http://iopscience.iop.org/0953-8984/16/15/R03.

[89] Peter J. Bond et al. “Coarse-grained molecular dynamics simulations of membrane proteins and peptides”. In: *Journal of Structural Biology* 157.3 (Mar. 2007), pp. 593–605. ISSN: 1047-8477. DOI: 10.1016/j.jsb.2006.10.004. URL: http://www.sciencedirect.com/science/article/pii/S1047847706003108.

[90] Siewert J Marrink et al. “The MARTINI force field: coarse grained model for biomolecular simulations”. eng. In: *The journal of physical chemistry. B* 111.27 (July 2007), pp. 7812–7824. ISSN: 1520-6106. DOI: 10.1021/jp071097f.

[91] Matthias R Schmidt et al. “Simulation-based prediction of phosphatidylinositol 4,5-bisphosphate binding to an ion channel”. eng. In: *Biochemistry* 52.2 (Jan. 2013), pp. 279–281. ISSN: 1520-4995. DOI: 10.1021/bi301350s.

[92] Werner Treptow, Siewert-J Marrink, and Mounir Tarek. “Gating Motions in Voltage-Gated Potassium Channels Revealed by Coarse-Grained Molecular Dynamics Simulations”. In: *The Journal of Physical Chemistry B* 112.11 (Mar. 2008), pp. 3277–3282. ISSN: 1520-6106. DOI: 10.1021/jp709675e. URL: http://dx.doi.org/10.1021/jp709675e.

[93] Guy G. Dodson, David P. Lane, and Chandra S. Verma. “Molecular simulations of protein dynamics: new windows on mechanisms in biology”. en. In: *EMBO reports* 9.2 (Feb. 2008), pp. 144–150. ISSN: 1469-221X. DOI: 10.1038/sj.embor.7401160. URL: http://www.nature.com/embor/journal/v9/n2/full/7401160.html.

[94] Ron O Dror et al. “Biomolecular simulation: a computational microscope for molecular biology”. eng. In: *Annual review of biophysics* 41 (2012), pp. 429–452. ISSN: 1936-1238. DOI: 10.1146/annurev-biophys-042910-155245.

[95] David E. Shaw et al. “Atomic-Level Characterization of the Structural Dynamics of Proteins”. en. In: *Science* 330.6002 (Oct. 2010), pp. 341–346. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1187409. URL: http://www.sciencemag.org/content/330/6002/341.

[96] Allard J Katan and Cees Dekker. “High-speed AFM reveals the dynamics of single biomolecules at the nanometer scale”. eng. In: *Cell* 147.5 (Nov. 2011), pp. 979–982. ISSN: 1097-4172. DOI: 10.1016/j.cell.2011.11.017.

[97] Cosmin I. Blaga et al. “Imaging ultrafast molecular dynamics with laser-induced electron diffraction”. en. In: *Nature* 483.7388 (Mar. 2012), pp. 194–197. ISSN: 0028-0836. DOI: 10.1038/nature10820. URL: http://www.nature.com/nature/journal/v483/n7388/full/nature10820.html.

[98] K. Binder. “Monte-Carlo Methods”. en. In: *Mathematical Tools for Physicists*. Ed. by George L. Trigg. Wiley-VCH Verlag GmbH & Co. KGaA, 2006, pp. 249–280. ISBN: 9783527607778. URL: http://onlinelibrary.wiley.com/doi/10.1002/3527607773.ch9/summary.

[99] Stewart A. Adcock and J. Andrew McCammon. “Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins”. In: *Chemical reviews* 106.5 (May 2006). ISSN: 0009-2665. DOI: 10.1021/cr040426m. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2547409/.

[100] Robert E. Bruccoleri and Martin Karplus. “Conformational sampling using high-temperature molecular dynamics”. en. In: *Biopolymers* 29.14 (1990), pp. 1847–1862. ISSN: 1097-0282. DOI: 10.1002/bip.360291415. URL: http://onlinelibrary.wiley.com/doi/10.1002/bip.360291415/abstract.

[101] Ryszard Czerminski and Ron Elber. “Computational studies of ligand diffusion in globins: I. Leghemoglobin”. en. In: *Proteins: Structure, Function, and Bioinformatics* 10.1 (1991), pp. 70–80. ISSN: 1097-0134. DOI: 10.1002/prot.340100107. URL: http://onlinelibrary.wiley.com/doi/10.1002/prot.340100107/abstract.

[102] René C. van Schaik et al. “A Structure Refinement Method Based on Molecular Dynamics in Four Spatial Dimensions”. In: *Journal of Molecular Biology* 234.3 (Dec. 1993), pp. 751–762. ISSN: 0022-2836. DOI: 10.1006/jmbi.1993.1624. URL: http://www.sciencedirect.com/science/article/pii/S0022283683716244.

[103] Peter Krüger et al. “Extending the capabilities of targeted molecular dynamics: Simulation of a large conformational transition in plasminogen activator inhibitor 1”. en. In: *Protein Science* 10.4 (2001), pp. 798–808. ISSN: 1469-896X. DOI: 10.1110/ps.40401. URL: http://onlinelibrary.wiley.com/doi/10.1110/ps.40401/abstract.

[104] Ayori Mitsutake, Yuji Sugita, and Yuko Okamoto. “Generalized-ensemble algorithms for molecular simulations of biopolymers”. en. In: *Peptide Science* 60.2 (2001), pp. 96–123. ISSN: 1097-0282. DOI: 10.1002/1097-0282(2001)60:2<96::AID-BIP1007>3.0.CO;2-F. URL: http://onlinelibrary.wiley.com/doi/10.1002/1097-0282(2001)60:2<96::AID-BIP1007>3.0.CO;2-F/abstract.

[105] K.y. Sanbonmatsu and A.e. García. “Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics”. en. In: *Proteins: Structure, Function, and Bioinformatics* 46.2 (2002), pp. 225–234. ISSN: 1097-0134. DOI: 10.1002/prot.1167. URL: http://onlinelibrary.wiley.com/doi/10.1002/prot.1167/abstract.

[106] Anton K. Faradjian and Ron Elber. “Computing time scales from reaction coordinates by milestoning”. In: *The Journal of Chemical Physics* 120.23 (June 2004), pp. 10880–10889. ISSN: 00219606. DOI: doi:10.1063/1.1738640. URL: http://jcp.aip.org/resource/1/jcpsa6/v120/i23/p10880_s1?isAuthorized=no.

[107] Berk Hess et al. “GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation”. In: *Journal of Chemical Theory and Computation* 4.3 (2008), pp. 435–447. ISSN: 1549-9618. DOI: 10.1021/ct700301q. URL: http://dx.doi.org/10.1021/ct700301q.

[108] Bernard R. Brooks et al. “CHARMM: A program for macromolecular energy, minimization, and dynamics calculations”. en. In: *Journal of Computational Chemistry* 4.2 (1983), pp. 187–217. ISSN: 1096-987X. DOI: 10.1002/jcc.540040211. URL: http://onlinelibrary.wiley.com/doi/10.1002/jcc.540040211/abstract.

[109] Wendy D. Cornell et al. “A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules”. In: *Journal of the American Chemical Society* 117.19 (May 1995), pp. 5179–5197. ISSN: 0002-7863. DOI: 10.1021/ja00124a002. URL: http://dx.doi.org/10.1021/ja00124a002.

[110] Mark T. Nelson et al. “NAMD: a Parallel, Object-Oriented Molecular Dynamics Program”. en. In: *International Journal of High Performance Computing Applications* 10.4 (Dec. 1996), pp. 251–268. ISSN: 1094-3420, 1741-2846. DOI: 10.1177/109434209601000401. URL: http://hpc.sagepub.com/content/10/4/251.

[111] Steve Plimpton. “Fast Parallel Algorithms for Short-Range Molecular Dynamics”. In: *Journal of Computational Physics* 117.1 (Mar. 1995), pp. 1–19. ISSN: 0021-9991. DOI: 10.1006/jcph.1995.1039. URL: http://www.sciencedirect.com/science/article/pii/S002199918571039X.

[112] Kresten Lindorff-Larsen et al. “How fast-folding proteins fold”. eng. In: *Science (New York, N.Y.)* 334.6055 (Oct. 2011), pp. 517–520. ISSN: 1095-9203. DOI: 10.1126/science.1208351.

[113] Stefano Piana, Kresten Lindorff-Larsen, and David E. Shaw. “Atomic-level description of ubiquitin folding”. en. In: *Proceedings of the National Academy of Sciences* (Mar. 2013). ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1218321110. URL: http://www.pnas.org/content/early/2013/03/14/1218321110.

[114] Aleksij Aksimentiev and Klaus Schulten. “Imaging alpha-hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map”. eng. In: *Biophysical journal* 88.6 (June 2005), pp. 3745–3761. ISSN: 0006-3495. DOI: 10.1529/biophysj.104.058727.

[115] Wolfgang Eckhardt et al. “591 TFLOPS Multi-trillion Particles Simulation on SuperMUC”. In: *Supercomputing*. Ed. by Julian Martin Kunkel, Thomas Ludwig, and Hans Werner Meuer. Lecture Notes in Computer Science 7905. Springer Berlin Heidelberg, Jan. 2013, pp. 1–12. ISBN: 978-3-642-38749-4, 978-3-642-38750-0. URL: http://link.springer.com/chapter/10.1007/978-3-642-38750-0_1.

[116] Swendsen and Wang. “Replica Monte Carlo simulation of spin glasses”. ENG. In: *Physical review letters* 57.21 (Nov. 1986), pp. 2607–2609. ISSN: 1079-7114.

[117] Ulrich H.E. Hansmann. “Parallel tempering algorithm for conformational studies of biological molecules”. In: *Chemical Physics Letters* 281.1-3 (Dec. 1997), pp. 140–150. ISSN: 0009-2614. DOI: 10.1016/S0009-2614(97)01198-6. URL: http://www.sciencedirect.com/science/article/pii/S0009261497011986.

[118] G.M. Torrie and J.P. Valleau. “Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling”. In: *Journal of Computational Physics* 23.2 (Feb. 1977), pp. 187–199. ISSN: 0021-9991. DOI: 10.1016/0021-9991(77)90121-8. URL: http://www.sciencedirect.com/science/article/pii/0021999177901218.

[119] Chris Oostenbrink et al. “A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6”. eng. In: *Journal of computational chemistry* 25.13 (Oct. 2004), pp. 1656–1676. ISSN: 0192-8651. DOI: 10.1002/jcc.20090.

[120] Lukas D. Schuler, Xavier Daura, and Wilfred F. van Gunsteren. “An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase”. en. In: *Journal of Computational Chemistry* 22.11 (2001), pp. 1205–1218. ISSN: 1096-987X. DOI: 10.1002/jcc.1078. URL: http://onlinelibrary.wiley.com/doi/10.1002/jcc.1078/abstract.

[121] Walter Ernest Reiher. *Theoretical Studies of Hydrogen Bonding*. en. Harvard University, 1985.

[122] MacKerell et al. “All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†”. In: *The Journal of Physical Chemistry B* 102.18 (Apr. 1998), pp. 3586–3616. ISSN: 1520-6106. DOI: 10.1021/jp973084f. URL: http://dx.doi.org/10.1021/jp973084f.

[123] Alexander D. Mackerell, Michael Feig, and Charles L. Brooks. “Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations”. en. In: *Journal of Computational Chemistry* 25.11 (2004), pp. 1400–1415. ISSN: 1096-987X. DOI: 10.1002/jcc.20065. URL: http://onlinelibrary.wiley.com/doi/10.1002/jcc.20065/abstract.

[124] Peter A. Kollman. “Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the Properties of Organic and Biological Molecules”. In: *Accounts of Chemical Research* 29.10 (Jan. 1996), pp. 461–469. ISSN: 0001-4842. DOI: 10.1021/ar9500675. URL: http://dx.doi.org/10.1021/ar9500675.

[125] 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?” en. In: *Journal of Computational Chemistry* 21.12 (2000), pp. 1049–1074. ISSN: 1096-987X. DOI: 10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F. URL: http://onlinelibrary.wiley.com/doi/10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F/abstract.

[126] Viktor Hornak et al. “Comparison of multiple Amber force fields and development of improved protein backbone parameters”. eng. In: *Proteins* 65.3 (Nov. 2006), pp. 712–725. ISSN: 1097-0134. DOI: 10.1002/prot.21123.

[127] Robert B. Best, Nicolae-Viorel Buchete, and Gerhard Hummer. “Are Current Molecular Dynamics Force Fields too Helical?” In: *Biophysical Journal* 95.1 (July 2008), pp. L07–L09. ISSN: 0006-3495. DOI: 10.1529/biophysj.108.132696. URL: http://www.sciencedirect.com/science/article/pii/S0006349508702777.

[128] Kresten Lindorff-Larsen et al. “Improved side-chain torsion potentials for the Amber ff99SB protein force field”. In: *Proteins* 78.8 (June 2010). ISSN: 0887-3585. DOI: 10.1002/prot.22711. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2970904/.

[129] T Herges and W Wenzel. “An all-atom force field for tertiary structure prediction of helical proteins”. eng. In: *Biophysical journal* 87.5 (Nov. 2004), pp. 3100–3109. ISSN: 0006-3495. DOI: 10.1529/biophysj.104.040071.

[130] Abhinav Verma and Wolfgang Wenzel. “A Free-Energy Approach for All-Atom Protein Simulation”. In: *Biophysical Journal* 96.9 (May 2009). ISSN: 0006-3495. DOI: 10.1016/j.bpj.2008.12.3921. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2711412/.

[131] William L. Jorgensen and Julian. Tirado-Rives. “The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin”. In: *Journal of the American Chemical Society* 110.6 (Mar. 1988), pp. 1657–1666. ISSN: 0002-7863. DOI: 10.1021/ja00214a001. URL: http://dx.doi.org/10.1021/ja00214a001.

[132] A. Einstein. “Die Grundlage der allgemeinen Relativitätstheorie”. In: *Annalen der Physik* 354 (1916), pp. 769–822. ISSN: 0003-3804. DOI: 10.1002/andp.19163540702. URL: http://adsabs.harvard.edu/abs/1916AnP...354..769E.

[133] J. E. Jones. “On the Determination of Molecular Fields. II. From the Equation of State of a Gas”. en. In: *Proceedings of the Royal Society of London. Series A* 106.738 (Oct. 1924), pp. 463–477. ISSN: 1364-5021, 1471-2946. DOI: 10.1098/rspa.1924.0082. URL: http://rspa.royalsocietypublishing.org/content/106/738/463.

[134] Philip M. Morse. “Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels”. In: *Physical Review* 34.1 (July 1929), pp. 57–64. DOI: 10.1103/PhysRev.34.57. URL: http://link.aps.org/doi/10.1103/PhysRev.34.57.

[135] Josh Barnes and Piet Hut. “A hierarchical O(N log N) force-calculation algorithm”. en. In: *Nature* 324.6096 (Dec. 1986), pp. 446–449. DOI: 10.1038/324446a0. URL: http://www.nature.com/nature/journal/v324/n6096/abs/324446a0.html.

[136] “TOP500 Supercomputer Site”. In: (). URL: http://www.top500.org.

[137] Jeffrey K. Noel et al. “SMOG@ctbp: simplified deployment of structure-based models in GROMACS”. In: *Nucl. Acids Res.* (June 2010), gkq498. DOI: 10.1093/nar/gkq498. URL: http://nar.oxfordjournals.org/cgi/content/abstract/gkq498v1.

[138] OpenMP Architecture Review Board. *OpenMP Application Program Interface Version 3.0*. May 2008. URL: http://www.openmp.org/mp-documents/spec30.pdf.

[139] OpenMP Architecture Review Board. *OpenMP Application Program Interface Version 4.0*. 2013. URL: http://www.openmp.org/mp-documents/OpenMP4.0.0.pdf.

[140] J de Keyzer, C van der Does, and A J M Driessen. “The bacterial translocase: a dynamic protein channel complex”. eng. In: *Cellular and molecular life sciences: CMLS* 60.10 (Oct. 2003), pp. 2034–2052. ISSN: 1420-682X. DOI: 10.1007/s00018-003-3006-y.

[141] Philip A. Lee, Danielle Tullman-Ercek, and George Georgiou. “The Bacterial Twin-Arginine Translocation Pathway”. In: *Annual Review of Microbiology* 60.1 (2006), pp. 373–395. DOI: 10.1146/annurev.micro.60.080805.142212. URL: http://www.annualreviews.org/doi/abs/10.1146/annurev.micro.60.080805.142212.

[142] Frank Sargent et al. “Sec-independent Protein Translocation in Escherichia coli A DISTINCT AND PIVOTAL ROLE FOR THE TatB PROTEIN”. en. In: *Journal of Biological Chemistry* 274.51 (Dec. 1999), pp. 36073–36082. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.274.51.36073. URL: http://www.jbc.org/content/274/51/36073.

[143] Albert Bolhuis et al. “TatB and TatC Form a Functional and Structural Unit of the Twin-arginine Translocase from Escherichia coli”. en. In: *Journal of Biological Chemistry* 276.23 (June 2001), pp. 20213–20219. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.M100682200. URL: http://www.jbc.org/content/276/23/20213.

[144] Ming-Ren Yen et al. “Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system”. eng. In: *Archives of microbiology* 177.6 (June 2002), pp. 441–450. ISSN: 0302-8933. DOI: 10.1007/s00203-002-0408-4.

[145] Julia Fröbel, Patrick Rose, and Matthias Müller. “Twin-arginine-dependent translocation of folded proteins”. eng. In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 367.1592 (Apr. 2012), pp. 1029–1046. ISSN: 1471-2970. DOI: 10.1098/rstb.2011.0202.

[146] H Mori and K Cline. “Post-translational protein translocation into thylakoids by the Sec and DeltapH-dependent pathways”. eng. In: *Biochimica et biophysica acta* 1541.1-2 (Dec. 2001), pp. 80–90. ISSN: 0006-3002.

[147] Tracy Palmer, Frank Sargent, and Ben C Berks. “Export of complex cofactor-containing proteins by the bacterial Tat pathway”. eng. In: *Trends in microbiology* 13.4 (Apr. 2005), pp. 175–180. ISSN: 0966-842X. DOI: 10.1016/j.tim.2005.02.002.

[148] Nathan N. Alder and Steven M. Theg. “Energetics of Protein Transport across Biological Membranes”. In: *Cell* 112.2 (Jan. 2003), pp. 231–242. ISSN: 0092-8674. DOI: 10.1016/S0092-8674(03)00032-1. URL: http://www.cell.com/abstract/S0092-8674(03)00032-1.

[149] Ben C Berks, Tracy Palmer, and Frank Sargent. “Protein targeting by the bacterial twin-arginine translocation (Tat) pathway”. In: *Current Opinion in Microbiology* 8.2 (Apr. 2005), pp. 174–181. ISSN: 1369-5274. DOI: 10.1016/j.mib.2005.02.010. URL: http://www.sciencedirect.com/science/article/pii/S1369527405000214.

[150] Tracy Palmer and Ben C Berks. “Moving folded proteins across the bacterial cell membrane”. eng. In: *Microbiology (Reading, England)* 149.Pt 3 (Mar. 2003), pp. 547–556. ISSN: 1350-0872.

[151] Ulrich Gohlke et al. “The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter”. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.30 (July 2005), pp. 10482–10486. ISSN: 0027-8424. DOI: 10.1073/pnas.0503558102. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1180781/.

[152] Yunfei Hu et al. “Solution NMR Structure of the TatA Component of the Twin-Arginine Protein Transport System from Gram-Positive Bacterium Bacillus subtilis”. In: *Journal of the American Chemical Society* 132.45 (Nov. 2010), pp. 15942–15944. ISSN: 0002-7863. DOI: 10.1021/ja1053785. URL: http://dx.doi.org/10.1021/ja1053785.

[153] S S Mande et al. “Protein-protein interactions in the pyruvate dehydrogenase multienzyme complex: dihydrolipoamide dehydrogenase complexed with the binding domain of dihydrolipoamide acetyltransferase”. eng. In: *Structure (London, England: 1993)* 4.3 (Mar. 1996), pp. 277–286. ISSN: 0969-2126.

[154] René A W Frank et al. “The molecular origins of specificity in the assembly of a multienzyme complex”. eng. In: *Structure (London, England: 1993)* 13.8 (Aug. 2005), pp. 1119–1130. ISSN: 0969-2126. DOI: 10.1016/j.str.2005.04.021.

[155] Kenneth J Rosenberg et al. “Complementary dimerization of microtubule-associated tau protein: Implications for microtubule bundling and tau-mediated pathogenesis”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 105.21 (May 2008), pp. 7445–7450. ISSN: 1091-6490. DOI: 10.1073/pnas.0802036105.

[156] Hyeongjun Kim et al. “Formation of salt bridges mediates internal dimerization of myosin VI medial tail domain”. eng. In: *Structure (London, England: 1993)* 18.11 (Nov. 2010), pp. 1443–1449. ISSN: 1878-4186. DOI: 10.1016/j.str.2010.09.011.

[157] Oliver Beckstein et al. “Zipping and unzipping of adenylate kinase: atomistic insights into the ensemble of open<–>closed transitions”. eng. In: *Journal of molecular biology* 394.1 (Nov. 2009), pp. 160–176. ISSN: 1089-8638. DOI: 10.1016/j.jmb.2009.09.009.

[158] UniProt Consortium. “Update on activities at the Universal Protein Resource (UniProt) in 2013”. eng. In: *Nucleic acids research* 41.Database issue (Jan. 2013), pp. D43–47. ISSN: 1362-4962. DOI: 10.1093/nar/gks1068.

[159] Paul C. Whitford et al. “An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields”. In: *Proteins: Structure, Function, and Bioinformatics* 75.2 (May 2009), pp. 430–441. ISSN: 08873585. DOI: 10.1002/prot.22253. URL: http://doi.wiley.com/10.1002/prot.22253.

[160] David Eisenberg and Andrew D. McLachlan. “Solvation energy in protein folding and binding”. In: *Nature* 319.6050 (Jan. 1986), pp. 199–203. DOI: 10.1038/319199a0. URL: http://dx.doi.org/10.1038/319199a0.

[161] Cecilia Clementi, Angel E García, and José N Onuchic. “Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: all-atom representation study of protein L”. eng. In: *Journal of molecular biology* 326.3 (Feb. 2003), pp. 933–954. ISSN: 0022-2836.

[162] Leslie L Chavez et al. “Multiple routes lead to the native state in the energy landscape of the beta-trefoil family”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 103.27 (July 2006), pp. 10254–10258. ISSN: 0027-8424. DOI: 10.1073/pnas.0510110103.

[163] Alexander Schug et al. “Mutations as trapdoors to two competing native conformations of the Rop-dimer”. eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 104.45 (Nov. 2007), pp. 17674–17679. ISSN: 0027-8424. DOI: 10.1073/pnas.0706077104.

[164] Alexander Schug et al. “High-resolution protein complexes from integrating genomic information with molecular simulation”. en. In: *Proceedings of the National Academy of Sciences* 106.52 (Dec. 2009), pp. 22124–22129. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0912100106. URL: http://www.pnas.org/content/106/52/22124.

[165] Massimiliano Bonomi et al. “PLUMED: A portable plugin for free-energy calculations with molecular dynamics”. In: *Computer Physics Communications* 180.10 (Oct. 2009), pp. 1961–1972. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2009.05.011. URL: http://www.sciencedirect.com/science/article/B6TJ5-4W91PWV-7/2/4b543dffb25a51c136ac931316269ef0.

[166] Christian Lange et al. “Structure analysis of the protein translocating channel TatA in membranes using a multi-construct approach”. eng. In: *Biochimica et biophysica acta* 1768.10 (Oct. 2007), pp. 2627–2634. ISSN: 0006-3002. DOI: 10.1016/j.bbamem.2007.06.021.

[167] Torsten H Walther et al. “Membrane alignment of the pore-forming component TatA(d) of the twin-arginine translocase from Bacillus subtilis resolved by solid-state NMR spectroscopy”. eng. In: *Journal of the American Chemical Society* 132.45 (Nov. 2010), pp. 15945–15956. ISSN: 1520-5126. DOI: 10.1021/ja106963s.

[168] B C Berks, F Sargent, and T Palmer. “The Tat protein export pathway”. eng. In: *Molecular microbiology* 35.2 (Jan. 2000), pp. 260–274. ISSN: 0950-382X.

[169] Mark C. Leake et al. “Variable stoichiometry of the TatA component of the twin-arginine protein transport system observed by in vivo single-molecule imaging”. en. In: *Proceedings of the National Academy of Sciences* 105.40 (Oct. 2008), pp. 15376–15381. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0806338105. URL: http://www.pnas.org/content/105/40/15376.

[170] Gemma Warren et al. “Contributions of the Transmembrane Domain and a Key Acidic Motif to Assembly and Function of the TatA Complex”. In: *Journal of Molecular Biology* 388.1 (Apr. 2009), pp. 122–132. ISSN: 0022-2836. DOI: 10.1016/j.jmb.2009.02.060. URL: http://www.sciencedirect.com/science/article/pii/S0022283609002423.

[171] Joanne Oates et al. “The Escherichia coli Twin-arginine Translocation Apparatus Incorporates a Distinct Form of TatABC Complex, Spectrum of Modular TatA Complexes and Minor TatAB Complex”. In: *Journal of Molecular Biology* 346.1 (Feb. 2005), pp. 295–305. ISSN: 0022-2836. DOI: 10.1016/j.jmb.2004.11.047. URL: http://www.sciencedirect.com/science/article/pii/S0022283604015062.

[172] Kamila Gouffi et al. “Dual Topology of the Escherichia coli TatA Protein”. en. In: *Journal of Biological Chemistry* 279.12 (Mar. 2004), pp. 11608–11615. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.M313187200. URL: http://www.jbc.org/content/279/12/11608.

[173] Carole Dabney-Smith, Hiroki Mori, and Kenneth Cline. “Requirement of a Tha4-conserved transmembrane glutamate in thylakoid Tat translocase assembly revealed by biochemical complementation”. eng. In: *The Journal of biological chemistry* 278.44 (Oct. 2003), pp. 43027–43033. ISSN: 0021-9258. DOI: 10.1074/jbc.M307923200.

[174] Catherine S Chan et al. “The TatA subunit of Escherichia coli twin-arginine translocase has an N-in topology”. eng. In: *Biochemistry* 46.25 (June 2007), pp. 7396–7404. ISSN: 0006-2960. DOI: 10.1021/bi7005288.

[175] Gaye F. White et al. “Subunit Organization in the TatA Complex of the Twin Arginine Protein Translocase A SITE-DIRECTED EPR SPIN LABELING STUDY”. en. In: *Journal of Biological Chemistry* 285.4 (Jan. 2010), pp. 2294–2301. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.M109.065458. URL: http://www.jbc.org/content/285/4/2294.

[176] Nora Vázquez-Laslop, Hyunwoo Lee, and Alexander A. Neyfakh. “Increased Persistence in Escherichia coli Caused by Controlled Expression of Toxins or Other Unrelated Proteins”. en. In: *Journal of Bacteriology* 188.10 (May 2006), pp. 3494–3497. ISSN: 0021-9193, 1098-5530. DOI: 10.1128/JB.188.10.3494-3497.2006. URL: http://jb.asm.org/content/188/10/3494.

[177] Kim Lewis. “Persister cells, dormancy and infectious disease”. en. In: *Nature Reviews Microbiology* 5.1 (Jan. 2007), pp. 48–56. ISSN: 1740-1526. DOI: 10.1038/nrmicro1557. URL: http://www.nature.com/nrmicro/journal/v5/n1/abs/nrmicro1557.html.

[178] Sonja Hansen, Kim Lewis, and Marin Vulić. “Role of Global Regulators and Nucleotide Metabolism in Antibiotic Tolerance in Escherichia coli”. en. In: *Antimicrobial Agents and Chemotherapy* 52.8 (Aug. 2008), pp. 2718–2726. ISSN: 0066-4804, 1098-6596. DOI: 10.1128/AAC.00144-08. URL: http://aac.asm.org/content/52/8/2718.

[179] Thomas Steinbrecher et al. “Peptide-Lipid Interactions of the Stress-Response Peptide TisB That Induces Bacterial Persistence”. In: *Biophysical Journal* 103.7 (Oct. 2012), pp. 1460–1469. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2012.07.060. URL: http://www.sciencedirect.com/science/article/pii/S0006349512009836.

[180] Cecilia Unoson and E. Gerhart H. Wagner. “A small SOS-induced toxin is targeted against the inner membrane in Escherichia coli”. en. In: *Molecular Microbiology* 70.1 (2008), pp. 258–270. ISSN: 1365-2958. DOI: 10.1111/j.1365-2958.2008.06416.x. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2958.2008.06416.x/abstract.

[181] Jörg Vogel et al. “The Small RNA IstR Inhibits Synthesis of an SOS-Induced Toxic Peptide”. In: *Current Biology* 14.24 (Dec. 2004), pp. 2271–2276. ISSN: 0960-9822. DOI: 10.1016/j.cub.2004.12.003. URL: http://www.sciencedirect.com/science/article/pii/S096098220400942X.

[182] Tobias Dörr, Marin Vulić, and Kim Lewis. “Ciprofloxacin Causes Persister Formation by Inducing the TisB toxin in Escherichia coli”. In: *PLoS Biol* 8.2 (Feb. 2010), e1000317. DOI: 10.1371/journal.pbio.1000317. URL: http://dx.doi.org/10.1371/journal.pbio.1000317.

[183] Clinton C. Dawson, Chaidan Intapa, and Mary Ann Jabra-Rizk. ““Persisters”: Survival at the Cellular Level”. In: *PLoS Pathog* 7.7 (July 2011), e1002121. DOI: 10.1371/journal.ppat.1002121. URL: http://dx.doi.org/10.1371/journal.ppat.1002121.

[184] Michael E. Selsted and Andre J. Ouellette. “Mammalian defensins in the antimicrobial immune response”. en. In: *Nature Immunology* 6.6 (June 2005), pp. 551–557. ISSN: 1529-2908. DOI: 10.1038/ni1206. URL: http://www.nature.com/ni/journal/v6/n6/full/ni1206.html.

[185] Birgit Schittek et al. “Dermcidin: a novel human antibiotic peptide secreted by sweat glands”. en. In: *Nature Immunology* 2.12 (Dec. 2001), pp. 1133–1137. ISSN: 1529-2908. DOI: 10.1038/ni732. URL: http://www.nature.com/ni/journal/v2/n12/abs/ni732.html.

[186] R L Gallo et al. “Syndecans, cell surface heparan sulfate proteoglycans, are induced by a proline-rich antimicrobial peptide from wounds.” In: *Proceedings of the National Academy of Sciences of the United States of America* 91.23 (Nov. 1994). ISSN: 0027-8424. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC45161/.

[187] J Harder et al. “A peptide antibiotic from human skin”. eng. In: *Nature* 387.6636 (June 1997), p. 861. ISSN: 0028-0836. DOI: 10.1038/43088.

[188] Andreas Peschel and Hans-Georg Sahl. “The co-evolution of host cationic antimicrobial peptides and microbial resistance”. eng. In: *Nature reviews. Microbiology* 4.7 (July 2006), pp. 529–536. ISSN: 1740-1526. DOI: 10.1038/nrmicro1441.

[189] Maren Paulmann et al. “Structure-Activity Analysis of the Dermcidin-derived Peptide DCD-1L, an Anionic Antimicrobial Peptide Present in Human Sweat”. en. In: *Journal of Biological Chemistry* 287.11 (Mar. 2012), pp. 8434–8443. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.M111.332270. URL: http://www.jbc.org/content/287/11/8434.

[190] Sandra Burrack et al. “A new type of intracellular retention signal identified in a pestivirus structural glycoprotein”. en. In: *The FASEB Journal* 26.8 (Aug. 2012), pp. 3292–3305. ISSN: 0892-6638, 1530-6860. DOI: 10.1096/fj.12-207191. URL: http://www.fasebj.org/content/26/8/3292.

[191] Birke Andrea Tews and Gregor Meyers. “The Pestivirus Glycoprotein Erns Is Anchored in Plane in the Membrane via an Amphipathic Helix”. en. In: *Journal of Biological Chemistry* 282.45 (Nov. 2007), pp. 32730–32741. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.M706803200. URL: http://www.jbc.org/content/282/45/32730.

[192] Brett D Lindenbach and Charles M Rice. “Unravelling hepatitis C virus replication from genome to function”. eng. In: *Nature* 436.7053 (Aug. 2005), pp. 933–938. ISSN: 1476-4687. DOI: 10.1038/nature04077.

[193] Hiroyuki Horiuchi et al. “Primary Structure of a Base Non-Specific Ribonuclease from Rhizopus niveus”. en. In: *Journal of Biochemistry* 103.3 (Mar. 1988), pp. 408–418. ISSN: 0021-924X, URL: http://jb.oxfordjournals.org/content/103/3/408.

[194] R. Schneider et al. “Identification of a structural glycoprotein of an RNA virus as a ribonuclease”. en. In: *Science* 261.5125 (Aug. 1993), pp. 1169–1171. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.8356450. URL: http://www.sciencemag.org/content/261/5125/1169.

[195] Marcel M. Hulst and Rob J.M. Moormann. “[35] Erns protein of pestiviruses”. In: *Methods in Enzymology*. Ed. by Allen W. Nicholson. Vol. Volume 342. Ribonucleases - Part B. Academic Press, 2001, pp. 431–440. ISBN: 0076-6879. URL: http://www.sciencedirect.com/science/article/pii/S007668790142564X.

[196] Marcel M. Hulst et al. “Glycoprotein E2 of Classical Swine Fever Virus: Expression in Insect Cells and Identification as a Ribonuclease”. In: *Virology* 200.2 (May 1994), pp. 558–565. ISSN: 0042-6822. DOI: 10.1006/viro.1994.1218. URL: http://www.sciencedirect.com/science/article/pii/S0042682284712189.

[197] T Rumenapf et al. “Processing of the envelope glycoproteins of pestiviruses.” In: *Journal of Virology* 67.6 (June 1993). ISSN: 0022-538X. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC237670/.

[198] Christiane Fetzer, Birke Andrea Tews, and Gregor Meyers. “The Carboxy-Terminal Sequence of the Pestivirus Glycoprotein Erns Represents an Unusual Type of Membrane Anchor”. en. In: *Journal of Virology* 79.18 (Sept. 2005), pp. 11901–11913. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.79.18.11901-11913.2005. URL: http://jvi.asm.org/content/79/18/11901.

[199] Munir Iqbal, Helen Flick-Smith, and John W. McCauley. “Interactions of bovine viral diarrhoea virus glycoprotein Erns with cell surface glycosaminoglycans”. en. In: *Journal of General Virology* 81.2 (Feb. 2000), pp. 451–459. ISSN: 0022-1317, 1465-2099. URL: http://vir.sgmjournals.org/content/81/2/451.

[200] M. M. Hulst, H. G. P. van Gennip, and R. J. M. Moormann. “Passage of Classical Swine Fever Virus in Cultured Swine Kidney Cells Selects Virus Variants That Bind to Heparan Sulfate due to a Single Amino Acid Change in Envelope Protein Erns”. en. In: *Journal of Virology* 74.20 (Oct. 2000), pp. 9553–9561. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.74.20.9553-9561.2000. URL: http://jvi.asm.org/content/74/20/9553.

[201] M. M. Hulst et al. “Interaction of Classical Swine Fever Virus with Membrane-Associated Heparan Sulfate: Role for Virus Replication In Vivo and Virulence”. en. In: *Journal of Virology* 75.20 (Oct. 2001), pp. 9585–9595. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.75.20.9585-9595.2001. URL: http://jvi.asm.org/content/75/20/9585.

[202] Johannes P. M. Langedijk. “Translocation Activity of C-terminal Domain of Pestivirus Erns and Ribotoxin L3 Loop”. en. In: *Journal of Biological Chemistry* 277.7 (Feb. 2002), pp. 5308–5314. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.M104147200. URL: http://www.jbc.org/content/277/7/5308.

[203] E Weiland et al. “Pestivirus glycoprotein which induces neutralizing antibodies forms part of a disulfide-linked heterodimer”. eng. In: *Journal of virology* 64.8 (Aug. 1990), pp. 3563–3569. ISSN: 0022-538X.

[204] M. N. Widjojoatmodjo et al. “Classical Swine Fever Virus ErnsDeletion Mutants: trans-Complementation and Potential Use as Nontransmissible, Modified, Live-Attenuated Marker Vaccines”. en. In: *Journal of Virology* 74.7 (Apr. 2000), pp. 2973–2980. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.74.7.2973-2980.2000. URL: http://jvi.asm.org/content/74/7/2973.

[205] Ilona Reimann, Ilia Semmler, and Martin Beer. “Packaged replicons of bovine viral diarrhea virus are capable of inducing a protective immune response”. In: *Virology* 366.2 (Sept. 2007), pp. 377–386. ISSN: 0042-6822. DOI: 10.1016/j.virol.2007.05.006. URL: http://www.sciencedirect.com/science/article/pii/S0042682207003455.

[206] Gregor Meyers, Armin Saalmüller, and Mathias Büttner. “Mutations Abrogating the RNase Activity in Glycoprotein Erns of the Pestivirus Classical Swine Fever Virus Lead to Virus Attenuation”. en. In: *Journal of Virology* 73.12 (Dec. 1999), pp. 10224–10235. ISSN: 0022-538X, 1098-5514. URL: http://jvi.asm.org/content/73/12/10224.

[207] Christiane Meyer et al. “Recovery of Virulent and RNase-Negative Attenuated Type 2 Bovine Viral Diarrhea Viruses from Infectious cDNA Clones”. en. In: *Journal of Virology* 76.16 (Aug. 2002), pp. 8494–8503. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.76.16.8494-8503.2002. URL: http://jvi.asm.org/content/76/16/8494.

[208] I. Fernandez Sainz et al. “Removal of a N-linked glycosylation site of classical swine fever virus strain Brescia Erns glycoprotein affects virulence in swine”. In: *Virology* 370.1 (Jan. 2008), pp. 122–129. ISSN: 0042-6822. DOI: 10.1016/j.virol.2007.08.028. URL: http://www.sciencedirect.com/science/article/pii/S0042682207005624.

[209] David Brown et al. “A domain decomposition parallelization strategy for molecular dynamics simulations on distributed memory machines”. In: *Computer Physics Communications* 74.1 (Jan. 1993), pp. 67–80. ISSN: 0010-4655. DOI: 10.1016/0010-4655(93)90107-N. URL: http://www.sciencedirect.com/science/article/pii/001046559390107N.

[210] Jun S. Liu, Faming Liang, and Wing Hung Wong. “The Multiple-Try Method and Local Optimization in Metropolis Sampling”. In: *Journal of the American Statistical Association* 95.449 (Mar. 2000), p. 121. ISSN: 01621459. DOI: 10.2307/2669532. URL: http://www.jstor.org/discover/10.2307/2669532?uid=3737864&uid=2&uid=4&sid=21102060368013.

[211] Ming-Hui Chen and Bruce Schmeiser. “Performance of the Gibbs, Hit-and-Run, and Metropolis Samplers”. In: *Journal of Computational and Graphical Statistics* 2.3 (Sept. 1993). ArticleType: research-article / Full publication date: Sep., 1993 / Copyright © 1993 American Statistical Association, Institute of Mathematical Statistics and Interface Foundation of America, pp. 251–272. ISSN: 1061-8600. DOI: 10.2307/1390645. URL: http://www.jstor.org/stable/1390645.

[212] Roy J. Glauber. “Time-Dependent Statistics of the Ising Model”. In: *Journal of Mathematical Physics* 4.2 (Feb. 1963), p. 294. ISSN: 00222488. DOI: doi:10.1063/1.1703954. URL: http://jmp.aip.org/resource/1/jmapaq/v4/i2/p294_s1.

[213] C. James McKnight, Paul T. Matsudaira, and Peter S. Kim. “NMR structure of the 35-residue villin headpiece subdomain”. In: *Nature Structural Biology* 4.3 (Mar. 1997), pp. 180–184. ISSN: 1072-8368. DOI: 10.1038/nsb0397-180. URL: http://europepmc.org/abstract/MED/9164455/reload=0;jsessionid=UU6w77OhE7g4zkzcW8gN.24.

[214] David E. Shaw et al. “Anton, a special-purpose machine for molecular dynamics simulation”. In: *Commun. ACM* 51.7 (July 2008), 91–97. ISSN: 0001-0782. DOI: 10.1145/1364782.1364802. URL: http://doi.acm.org/10.1145/1364782.1364802.

[215] J.S. Kuskin et al. “Incorporating flexibility in Anton, a specialized machine for molecular dynamics simulation”. In: *IEEE 14th International Symposium on High Performance Computer Architecture, 2008. HPCA 2008*. 2008, pp. 343–354. DOI: 10.1109/HPCA.2008.4658651.

[216] David E. Shaw et al. “Millisecond-scale molecular dynamics simulations on Anton”. In: *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*. SC ’09. New York, NY, USA: ACM, 2009, 65:1–65:11. ISBN: 978-1-60558-744-8. DOI: 10.1145/1654059.1654126. URL: http://doi.acm.org/10.1145/1654059.1654126.

[217] David Shortle. “One sequence plus one mutation equals two folds”. In: *Proceedings of the National Academy of Sciences of the United States of America* 106.50 (Dec. 2009), pp. 21011–21012. ISSN: 0027-8424. DOI: 10.1073/pnas.0912370107. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2795560/.

[218] Tammie L. S. Benzinger et al. “Propagating structure of Alzheimer’s beta-amyloid(10–35) is parallel beta-sheet with residues in exact register”. en. In: *Proceedings of the National Academy of Sciences* 95.23 (Nov. 1998), pp. 13407–13412. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.95.23.13407. URL: http://www.pnas.org/content/95/23/13407.

[219] Abhinav Verma et al. “All-atom de novo protein folding with a scalable evolutionary algorithm”. en. In: *Journal of Computational Chemistry* 28.16 (2007), pp. 2552–2558. ISSN: 1096-987X. DOI: 10.1002/jcc.20750. URL: http://onlinelibrary.wiley.com/doi/10.1002/jcc.20750/abstract.

[220] Robert B Best and Gerhard Hummer. “Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides”. eng. In: *The journal of physical chemistry. B* 113.26 (July 2009), pp. 9004–9015. ISSN: 1520-6106. DOI: 10.1021/jp901540t.

[221] Stefano Piana, Kresten Lindorff-Larsen, and David E. Shaw. “Protein folding kinetics and thermodynamics from atomistic simulation”. en. In: *Proceedings of the National Academy of Sciences* 109.44 (Oct. 2012), pp. 17845–17850. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1201811109. URL: http://www.pnas.org/content/109/44/17845.

[222] Jeffrey K. Noel, Paul C. Whitford, and José N. Onuchic. “The Shadow Map: A General Contact Definition for Capturing the Dynamics of Biomolecular Folding and Function”. In: *The Journal of Physical Chemistry B* 116.29 (July 2012), pp. 8692–8702. ISSN: 1520-6106. DOI: 10.1021/jp300852d. URL: http://dx.doi.org/10.1021/jp300852d.

[223] William Moffitt and Jen Tsi Yang. “THE OPTICAL ROTATORY DISPERSION OF SIMPLE POLYPEPTIDES. I”. In: *Proceedings of the National Academy of Sciences of the United States of America* 42.9 (Sept. 1956). ISSN: 0027-8424. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC534258/.

[224] William Moffitt. “THE OPTICAL ROTATORY DISPERSION OF SIMPLE POLYPEPTIDES. II”. In: *Proceedings of the National Academy of Sciences of the United States of America* 42.10 (Oct. 1956). ISSN: 0027-8424. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC528325/.

[225] William Moffitt. “Optical Rotatory Dispersion of Helical Polymers”. In: *The Journal of Chemical Physics* 25.3 (Sept. 1956), pp. 467–478. ISSN: 00219606. DOI: doi:10.1063/1.1742946. URL: http://jcp.aip.org/resource/1/jcpsa6/v25/i3/p467_s1?isAuthorized=no.

[226] Benjamin M. Bulheller, Alison Rodger, and Jonathan D. Hirst. “Circular and linear dichroism of proteins”. en. In: *Physical Chemistry Chemical Physics* 9.17 (Apr. 2007), pp. 2020–2035. ISSN: 1463-9084. DOI: 10.1039/B615870F. URL: http://pubs.rsc.org/en/content/articlelanding/2007/cp/b615870f.

[227] Jonathan D. Hirst, Samita Bhattacharjee, and Alexey V. Onufriev. “Theoretical studies of time-resolved spectroscopy of protein folding”. en. In: *Faraday Discussions* 122.0 (Oct. 2003), pp. 253–267. ISSN: 1364-5498. DOI: 10.1039/B200714B. URL: http://pubs.rsc.org/en/content/articlelanding/2003/fd/b200714b.

[228] Robert W. Woody. “[4] Circular dichroism”. In: *Methods in Enzymology*. Ed. by Kenneth Sauer. Vol. Volume 246. Biochemical Spectroscopy. Academic Press, 1995, pp. 34–71. ISBN: 0076-6879. URL: http://www.sciencedirect.com/science/article/pii/0076687995460063.

[229] Minghui Wang et al. “Dynamic NMR Line-Shape Analysis Demonstrates that the Villin Headpiece Subdomain Folds on the Microsecond Time Scale”. In: *Journal of the American Chemical Society* 125.20 (May 2003), pp. 6032–6033. ISSN: 0002-7863. DOI: 10.1021/ja028752b. URL: http://dx.doi.org/10.1021/ja028752b.

[230] Yong Duan and Peter A. Kollman. “Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution”. en. In: *Science* 282.5389 (Oct. 1998), pp. 740–744. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.282.5389.740. URL: http://www.sciencemag.org/content/282/5389/740.

[231] Samuel S. Cho, Yaakov Levy, and Peter G. Wolynes. “Quantitative criteria for native energetic heterogeneity influences in the prediction of protein folding kinetics”. en. In: *Proceedings of the National Academy of Sciences* 106.2 (Jan. 2009), pp. 434–439. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0810218105. URL: http://www.pnas.org/content/106/2/434.

[232] Alan R. Fersht and Satoshi Sato. “Phi-Value analysis and the nature of protein-folding transition states”. In: *Proceedings of the National Academy of Sciences of the United States of America* 101.21 (May 2004), pp. 7976–7981. ISSN: 0027-8424. DOI: 10.1073/pnas.0402684101. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC419542/.

[233] Jan Kubelka et al. “Chemical, physical, and theoretical kinetics of an ultrafast folding protein”. en. In: *Proceedings of the National Academy of Sciences* 105.48 (Dec. 2008), pp. 18655–18662. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0808600105. URL: http://www.pnas.org/content/105/48/18655.

[234] Shankar Kumar et al. “THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method”. en. In: *Journal of Computational Chemistry* 13.8 (1992), pp. 1011–1021. ISSN: 1096-987X. DOI: 10.1002/jcc.540130812. URL: http://onlinelibrary.wiley.com/doi/10.1002/jcc.540130812/abstract.

[235] Jeffrey K. Noel et al. “SMOG@ctbp: simplified deployment of structure-based models in GROMACS”. en. In: *Nucleic Acids Research* 38.suppl 2 (July 2010), W657–W661. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkq498. URL: http://nar.oxfordjournals.org/content/38/suppl_2/W657.

[236] Emilio Gallicchio, Kristina Paris, and Ronald M. Levy. “The AGBNP2 Implicit Solvation Model”. In: *Journal of Chemical Theory and Computation* 5.9 (Sept. 2009), pp. 2544–2564. ISSN: 1549-9618. DOI: 10.1021/ct900234u. URL: http://dx.doi.org/10.1021/ct900234u.

Acknowledgments

First and foremost, I want to thank the Baden Württemberg Stiftung (foundation of the State of Baden-Wuerttemberg) for funding my thesis within the HPC-5 project: *High Throughput Protein Structure Prediction on Hybrid and Distributed High Performance Architectures*.
In particular, I wish to express my gratitude to my supervisor Prof. Dr. Wolfgang Wenzel for his continued encouragement and invaluable suggestions during this work. In this I would also like to include my gratitude to Prof. Dr. Ulrich Nienhaus for refereeing my thesis.
I would like to thank all my collaboration partners, in particular RG Ulrich (KIT) and RG Ruggerone (Università di Cagliari) for the enjoyable cooperation in the investigation of the TatA complex and Alexander Schug and his group (KIT) for exciting protein studies.
I want to thank all my colleagues, especially those in my office: Timo Strunk, who initiated the SIMONA project and supplied me with many useful advices over my time at the INT. Without this great teamwork, the SIMONA project would not be where it is now. Nana Phoenix Heilmann for her relentless fight against the Expresso 2000 bean and Thomas Cooks for his great work in public relations for the poem@home project. Special thanks go to our head cook Julia Setzler, for her dedication to our cooking group and to Martin Brieg for his valuable contribution to SIMONA, developing efficient GB approaches. Sorted by increasing office-to-office distance I would like to thank all the other members of the RG Wenzel and RG Schug: Priya Anand, Konstantin Klenin, Denis Danilov, Paul Kleine, Benedikt Schönauer, Frank Tristram, Simon Widmaier, Tobias Neumann, Velimir Meded, Franz Symalla, Igor Beljakov, Pascal Friedrich, Monika Borkowska-Panek, Angela Poschlad, Claude Sinner and Benjamin Lutz and former members: Robert Maul, Horacio Pérez-Sánchez, Irene Meliciani, Alexander Biewer, Felix Ehrler and Carolin Seith, Louisa Scholz for the good working atmosphere, fruitful discussions, numerous conferences and for replacing some birthday cakes by pretzels.
I thank my whole family for supporting me during my thesis. I would not have made it this far without them. Finally, I would like to thank my wife Eva for the wonderful time we have together and for encouraging me during my thesis. She is the real reason, why I implemented **EV**olutionary **A**lgorithms in SIMONA.

Publications

1. Structure of the Membrane Anchor of Pestivirus Glycoprotein E^{rns} , a Long Tilted Amphipathic Helix
   Daniel Aberle, Claudia Muhle-Goll, Jochen Bürck, ***Moritz Wolf****, Sabine Reißer, Burkhard Luy, Wolfgang Wenzel, Anne S. Ulrich, Gregor Meyers*
   PLoS Pathog 10, (2014)

2. Folding and Self-Assembly of the TatA Translocation Pore Based on a Charge Zipper Mechanism
   Torsten H. Walther, Christina Gottselig, Stephan L. Grage, ***Moritz Wolf****, Attilio V. Vargiu, Marco J. Klein, Stefanie Vollmer, Sebastian Prock, Mareike Hartmann, Sergiy Afonin, Eva Stockwald, Hartmut Heinzmann, Olga V. Nolandt, Wolfgang Wenzel, Paolo Ruggerone, Anne S. Ulrich*
   Cell Volume 152, Issues 1–2, 17 (2013), Pages 316–326

3. SIMONA 1.0: an efficient and versatile framework for stochastic simulations of molecular and nanoscale systems
   Moritz Wolf*, Timo Strunk, Martin Brieg, Konstantin Klenin, Alexander Biewer, Frank Tristram, Matthias Ernst, Paul-Jakob Kleine, Nana Heilmann, Ivan Kondov, Wolfgang Wenzel*
   Journal of computational chemistry 33.32 (2012)

4. Peptide structure prediction using distributed volunteer computing networks
   Timo Strunk, ***Moritz Wolf****, Wolfgang Wenzel*
   Journal of Mathematical Chemistry 50, no. 2 (2012): 421-428.

5. Proteinfaltung mittels Kinetic Monte Carlo Algorithmus
   Moritz Wolf
   Diploma thesis, Universität Karlsruhe (TH), (29. June 2010).

6. Complete Computational Characterization of Protein Folding Equilibrium at the All Atom level.
   Moritz Wolf*, Nana Heilmann, Wolfgang Wenzel*
   (in preparation)

7. Accelerated Monte-Carlo Simulations for All-Atom Protein Folding.
   Moritz Wolf*, Timo Strunk, Wolfgang Wenzel*
   (in preparation)

Proceedings

1. Benchmarking the POEM@HOME Network for Protein Structure Prediction
   Timo Strunk, Priya Anand, Martin Brieg, ***Moritz Wolf****, Konstantin Klenin, Irene Meliciani, Frank Tristram, Ivan Kondov, Wolfgang Wenzel*
   Proceedings of the 3rd International Workshop on Science Gateways for Life Sciences (2011)

2. Development and evaluation of a GPU-optimized N-body term for the simulation of biomolecules
   Timo Strunk, ***Moritz Wolf****, Wolfgang Wenzel*
   Computational Methods in Science and Engineering, Proceedings of the Workshop SimLabs@KIT, (2010) p. 35.

Biophysical Society Meeting Abstracts

1. Performance of An All-Atom Free Energy Approach For Protein Structure Prediction
   Anand, Priya, Strunk, Timo, Brieg, Martin, Meliciani, Irene, ***Wolf, Moritz****, Klenin, Konstantin, Wenzel, Wolfgang*
   Biophysical Journal 100, 48a (2011)

2. Folding and Self-Assembly of the Pore-Forming Unit Tat-A of the Bacterial Twin-Arginine Translocase
   Grage, Stephan L., Walther, Torsten H., ***Wolf, Moritz****, Vargiu, Attilio, Klein, Marco J., Ruggerone, Paolo, Wenzel, Wolfgang, Ulrich, Anne S.*
   Biophysical Journal 100, 345a (2011)

3. Analysis of Amino Acid Specific Energy Contributions to Native Conformations in High-Resolution Protein Structures
   Strunk, Timo, ***Wolf, Moritz****, Wenzel, Wolfgang*
   Biophysical Journal 102, 456a (2012)

4. Thermodynamic Characterization of Protein Folding Equilibriums at the All Atom Level
   Heilmann, Nana, Setzler, Julia, Brieg, Martin, Strunk, Timo, ***Wolf, Moritz****, Seith, Carolin, Wenzel, Wolfgang*
   Biophysical Journal 104, 369a-370a (2013)

5. Absolute Quality Assessment of Protein Models
   Strunk, Timo, ***Wolf, Moritz****, Wenzel, Wolfgang*
   Biophysical Journal 104, 229a (2013)

6. Modeling Assembly of the Tata Pore Forming Complex using an Implicit Membrane Model
   Wolf, Moritz*, Walther, Torsten H., Gottselig, Christina, Grage, Stephan L., Vargiu, Attilio, Klein, Marco J., Vollmer, Stefanie, Prock, Sebastian, Hartmann, Mareike, Afonin, Sergiy, Stockwald, Eva, Heinzmann, Hartmut, Wenzel, Wolfgang, Ruggerone, Paolo, Ulrich, Anne S.*
   Biophysical Journal 104, 288a (2013)

7. Accelerated Monte-Carlo Simulations for All-Atom Protein Folding
   Wolf, Moritz*, Strunk, Timo, Wenzel, Wolfgang*
   Biophysical Journal 106.2, 260a (2014)

Invited Talks

1. Simulating the TatA complex in an implicit hydropathy-based pore potential by modifying Gromacs
   ***Moritz Wolf**, Torsten H. Walther, Christina Gottselig, Stephan L. Grage, Attilio V. Vargiu, Marco J. Klein, Stefanie Vollmer, Sebastian Prock, Mareike Hartmann, Sergiy Afonin, Eva Stockwald, Hartmut Heinzmann, Olga V. Nolandt, Paolo Ruggerone, Anne S. Ulrich, Wolfgang Wenzel*
   RG Ulrich Group Meeting in Freudenstadt (2011)

Talks

1. High Throughput Protein Structure Prediction on Hybrid and Distributed High Performance Architectures
   ***Moritz Wolf**, Timo Strunk, Martin Brieg, Ivan Kondov, Wolfgang Wenzel*
   High Performance Computing in Science and Engineering, Stuttgart, 2013

2. All-Atom Modeling of Protein Folding and Aggregation
   ***Moritz Wolf**, Torsten H. Walther, Christina Gottselig, Stephan L. Grage, Attilio V. Vargiu, Marco J. Klein, Stefanie Vollmer, Sebastian Prock, Mareike Hartmann, Sergiy Afonin, Eva Stockwald, Hartmut Heinzmann, Olga V. Nolandt, Paolo Ruggerone, Anne S. Ulrich, Wolfgang Wenzel*
   Computer Simulation and Theory of Macromolecules 2013, Hünfeld, April 26-27, 2013