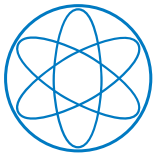
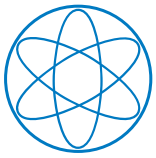


TECHNISCHE UNIVERSITÄT
MÜNCHEN
Physik Department
Lehrstuhl für Molekulardynamik T38

Dissertation

Advanced sampling to study protein plasticity and protein association

Katja Ostermeir



TECHNISCHE UNIVERSITÄT
MÜNCHEN
Physik Department
Lehrstuhl für Molekulardynamik T38

Advanced sampling to study protein plasticity and protein association

Dipl.-Phys. Univ.
Katja Ostermeir

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. Friedrich Simmel
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. Martin Zacharias
	2.	Univ.-Prof. Dr. Iris Antes (schriftliche Beurteilung)
		Univ.-Prof. Dr. Ulrich Gerland (mündliche Prüfung)

Die Dissertation wurde am 20.08.2014 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 17.10.2014 angenommen.

Abstract

Studying peptides and proteins by means of molecular dynamics simulations can give insights into their conformational variety on the atomistic scale. However, sampling is restricted by current computing capacities and the accuracy of force fields. If trapped in deep minima in a rough energetic landscape many proteins will rarely escape to cover the desired broad conformational range on a micro- to millisecond scale. Advanced sampling techniques such as Hamiltonian replica exchange (H-REMD) simulations aim at crossing energy barriers to enhance the conformational search for proteins and peptides in a set of parallel simulations. This thesis centers around the design of new efficient advanced sampling methods mainly involving H-REMD. A first approach makes use of specific biasing potentials to promote characteristic motions of these types of biomolecules. Backbone and side chain flexibility increases by penalizing characteristic dihedral angle combinations and side chain switches in neighboring replicas. Successfully capturing the folding process of peptides, this biasing potential replica exchange method (BP-REMD) also proves beneficial for the refinement of loop structures. Moreover, BP-REMD simulations demonstrated rotameric states of a side chain to be coupled to global binding pocket flexibility of MHC class I molecules. A second method is based on coarse grained elastic network models, which capture global motions of proteins by identifying their preferred directions of motion. This information is incorporated into atomistic H-REMD simulations in explicit solvent to account for large scale domain motions. Observed conformational changes of two test proteins affected their activity and function. Moreover, in both replica exchange methods, an adjustment protocol was designed to control the strength of biasing potentials to optimally exploit the replica framework.
Protein peptide interactions induce various biological processes and play a role in several diseases. Detailed information about binding free energy and binding poses on an atomistic level contribute to a better understanding. In this context, the H-REMD scheme improved binding free energy simulations that calculate free energy changes associated with alchemical transitions. Based on free energy perturbation theory, Lennard-Jones, Coulomb parameters and Born radii are transformed gradually in a single-topology series. In combination with the generalized Born implicit solvent model the relative binding free energy of modified ligands can be calculated rapidly.
Fragment based approaches aim at creating binders for molecular targets by first identifying low affinity ligands and their binding sites. Chemical assembly may then result in high affinity ligands for the desired target molecule. Inspired by this approach, dynamic self (re-)arrangement of molecule fragments is mimicked in MD simulations. In explicit solvent these fragments can combine into realistic molecules, which can degrade

and (re-)arrange dynamically during a simulation, to finally identify the best binder and a range of binding poses. The limits and potential of this method will be discussed. The advanced sampling methods developed in this thesis resulted in significantly enhanced sampling of local and global protein motion and rapid free energy calculations of protein complexes, allowing for a broad range of future applications.

Zusammenfassung

Molekulardynamiksimulationen auf atomistischer Skala geben Einblicke in die Dynamik von Peptiden und Proteinen. Die daraus resultierende Vielfalt der Konformationen ist durch derzeitige Rechenleistung und Genauigkeit der Kraftfelder eingeschränkt. Da viele Proteine auf der Mikro- bis Millisekundenskala in den tiefen Minima einer schroffen Energielandschaft gefangen sind, ändern sie ihre Form kaum. Fortgeschrittene Simulationstechniken wie Hamiltonian Replica Exchange Methoden (H-REMD) zielen darauf ab, Energiebarrieren zu überwinden, um die Konformationssuche für Peptide und Proteine in einem Set paralleler Simulationen zu beschleunigen. Im Zentrum dieser Arbeit steht die Entwicklung neuer Simulationsmethoden. Eine erste Methode greift auf Strafpotentiale zurück, die charakteristische Bewegungen dieser Biomoleküle forcieren. Die Flexibilität des Polymerrückgrats und der Seitenketten steigt dadurch, dass charakteristische Torsions- und Seitenkettenwinkel beeinflusst werden. Diese Biasing Potential H-REMD Methode (BP-REMD) beschleunigt ebenso die Faltung denaturierter Proteine wie deformierter Loop Strukturen. Darüberhinaus zeigte BP-REMD Methode, dass die Position einer Seitenkette die globale Flexibilität von MHC Klasse I Molekülen beeinflusst. Eine zweite Methode basiert auf grobkörnigen elastischen Netzwerk Modellen, die globale Proteinbewegungen beschreiben und bevorzugte Bewegungsrichtungen von Proteindomänen identifizieren. Integriert wurde diese Information in atomistische H-REMD Simulationen mit explizitem Wassermolekülen, um globale Bewegungen starrer Proteindomänen zu aktivieren. So waren an zwei Testproteinen Konformationsänderungen zu beobachten, die deren Funktion beeinflussen.
Ein Protokoll, das die Stärke der Strafpotentiale kontrolliert, wurde zusätzlich entwickelt, um das H-REMD Gerüst optimal auszunutzen. Protein-Peptid Interaktionen induzieren zahlreiche biologische Prozesse und spielen eine Rolle bei diversen Krankheiten. Detaillierte Informationen über freie Bindungsenergien und Bindepositionen auf atomistischem Niveau tragen insgesamt zu besserem Verständnis bei. In diesem Zusammenhang beschleunigte das H-REMD Schema die Vorhersage freier Bindeenergien: Schrittweise wurden in Störungsrechnungen Lennard-Jones und Coulomb Parameter sowie Born Radien in parallelen Simulationen umgewandelt. In Kombination mit dem Born Modell können Bindeenergien modifizierter Liganden schnell und effizient berechnet werden. Fragment basierte Methoden zielen darauf ab, hoch affine Liganden für bestimmte molekulare Targets zu kreieren: Zunächst werden Bindepositionen schwach bindender Liganden ermittelt. Verbindet man diese Fragmente chemisch, können hoch affine Liganden entstehen. Inspiriert von dieser Methode, wird dynamische Umordnung in MD Simulationen immitiert. In wässriger Lösung können sich schwach bindende Fragmente zu realistischen hoch affinen Liganden verbinden und eine Reihe

von möglichen Bindepositionen einnehmen. Grenzen und Möglichkeiten dieser Methode werden diskutiert.
Die in dieser Arbeit entwickelten fortgeschrittenen Simulationstechniken beschleunigten die Simulation globaler und lokaler Proteinbewegungen und freie Energierechnungen für Proteinkomplexe. Zugleich bieten sie ein breites Spektum für zukünftige Anwendungen.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank Prof. Dr. Martin Zacharias for his excellent guidance during my Ph.D. studies. Thank you for sharing your ideas on biomolecules, for involving me in this variety of diverse projects, for giving me so many insights how to investigate proteins with MD simulations and for sharpening my eyes for their dynamics. It was your open door, that made so many spontaneous discussions possible. I'm very grateful for everything I have learned, for your constant support and encouragement.
I would like to thank Prof. Dr. Iris Antes for assessing this thesis and Prof. Dr. Friedrich Simmel for chairing the examination committee.
I owe many thanks to Jill Seidlitz and Sonja Ortner for their help and support.
I would like to thank Prof. Dr. Sebastian Springer and Ph.D. Sunil Kumar Saini for introducing me into MHC molecules, for the fruitful collaboration and your hospitality.
I would like thank to the the group members of T38 for the good working atmosphere as well as for their helpfulness. Thanks to my roommates Sebastian and Alex for the good times. I'm grateful to Sebastian for proofreading.
Thanks to Kai and Antje for providing this template! A heartfelt thank you to my friends in and out of Munich! I'm especially grateful to Julia and Rasmus for being part of my life for so many years.
Many many thanks go to my parents and to my siblings Lisa and Felix for always being there for me and supporting and encouraging me in so many ways.

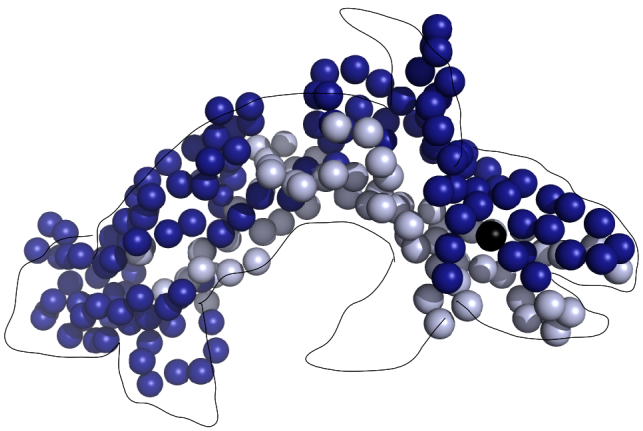


Figure 1: gpw130-D2-D3?

Chapter 1

Introduction

Proteins are of fundamental importance in living organisms. Encoded on genes in the cell nucleus, information on proteins is transcribed into messenger RNA for transport into cytoplasm. Finally synthesized by ribosomes, proteins perform various essential functions within every single cell: While cytoskeletal filaments can give shape to a cell, the proteasome complex degrades proteins to regulate their lifetime. Other proteins play an important role in cell communication and signaling. Transmembrane proteins permit specific substances to pass cell membranes to regulate, for instance, the communication between nerve cells. MHC class I molecules signal T-cells to destroy virus-infected cells: By transporting viral peptides from cytosol to cell surface, these protein complexes induce the viral immune response. Other types of proteins, hormones, spread in whole bloodstream to convey information from glands to their target organ to regulate physiological processes.
Usually, the protein function is coupled to a well defined three dimensional shape, the "native" state: Trapped in a rough energy landscape, randomly coiled proteins follow well defined folding pathways along a "folding funnel" on their way to functional form [110]. The latter is–according to Anfinsen's dogma–uniquely defined by their amino acid sequence [4]. However, this dogma is challenged by intrinsically disordered proteins: Flexible loop regions may connect well defined domains within a protein to allow for enhanced flexibility. Some proteins adopt their three dimensional structure only while binding to their target molecule, having several energetically comparable intermediates [44].
Studying proteins on an atomistic scale is not only an essential prerequisite for exploring protein function and protein-ligand interactions but also provides a basis for protein design. Thus, many experimental and computational efforts aim at elucidating conformations of biomolecules and investigating protein association. Experimental techniques as X-Ray crystallography [41] and nuclear magnetic resonance spectroscopy of proteins(NMR) [226] provide a static picture of a molecule at high atomistic resolution of the native state and sometimes even of intermediates. The dynamics of atoms of small peptides can be illuminated by NMR spectroscopy [8] or fast spectroscopic methods such as Förster resonance energy transfer (FRET) [177], however, usually at the expense of atomic and spatial resolution. Classical molecular dynamics (MD) simulations can bridge this gap by giving insights into the dynamics of biomolecules at atomistic

precision and at high time resolution [92]. Protein dynamics is driven by inter- and intramolecular forces respecting molecular geometry, electrostatic and Lennard-Jones interactions. Moreover, solvent-solute interactions are decisive for structure formation also including the hydrophobic effect [28]: While non-polar amino acids often form the stable protein core, hydrophilic ones are exposed to solvent.
However, classical MD simulations are limited by sampling time scales in the nano- to microsecond range and the accuracy of current force fields [161]. Thus, the first part of this thesis centers around the development of Hamiltonian replica exchange methods which focus on specifically accelerating conformational sampling of proteins.
Detailed information about interactions between proteins and peptides or other small molecules is essential for drug design or development of chemical compounds, which provide a basis for experimental studies. As key variable the binding free energy quantifies the stability of protein-peptide complexes. Hence, many efforts point in the direction to develop computational methods that calculate binding free energies to give insights into binding poses and binding kinetics. Docking approaches intend to predict binding poses of ligands on their target molecules in order to evaluate their binding affinity by scoring functions that respect geometrical restraints and energetic contributions [97, 49, 155]. However, these computationally efficient approaches do not consider protein flexibility and explicit water molecules: Ligand poses may deviate from experimentally observed positions and docking approaches may fail to classify binding affinity of ligands. Computational effort drastically increases in MD simulations, but involves high accuracy in a dynamical aqueous environment. Thus, the second part of this thesis deals with advanced sampling methods, which aim at elucidating experimental findings on dipeptide binding on MHC class I molecules, improving relative binding free energy calculations, and enabling dynamic self rearrangement of biomolecules.
This thesis is structured as follows:

- Chapter 2 addresses the essentials on proteins as well as on computational techniques, this thesis is based on: Dynamics of proteins can be successfully investigated by means of classical molecular dynamics simulations, which were developed into advanced sampling techniques as replica exchange molecular dynamics. Moreover, simulations have proven as effective tool to predict conformational free energy changes, binding free energies of ligand protein complexes and free energy changes of molecular conversions. Chapter 2 closes with an overview on free energy calculation approaches that are employed throughout the thesis.

- A Hamiltonian replica exchange method was designed to specifically promote local dihedral angle transitions along the polymer backbone and amino acid side chains flips. As discussed in chapter 3 protein folding and loop refinement accelerated compared to classical MD simulations.

- Domain-domain motions of proteins often are decisive for their functionality. In chapter 4 information about these global large scale motions is extracted from an elastic network model and integrated into a Hamiltonian replica exchange framework, to enhance conformational transitions of multi-domain proteins.

- MHC class I molecules are triggering the viral human response by presenting viral peptides to the cell surface to induce destruction of virus infected cells by

cytotoxic T-cells. Dynamics of two MHC class I allotypes is in the focus of chapter 5. Differing in one single amino acid (116) only, both allotypes bind peptides with different efficiency. MD simulations[†] showed allotypes to differ in their flexibility of the MHC class I binding pocket and suggested a molecular switch whose position is mediated by amino acid 116: The shape of the binding pocket turns out to be coupled to the position of the switch.

- High affinity peptides (consisting of 8-10 amino acids) are known to speed up folding of MHC class I molecules in their peptide receptive form in vitro. Experiments that were performed by Prof. Dr. Sebastian Springer and Ph.D. Sunil Kumar Saini prove dipeptides to fulfill the minimal requirements to enhance the folding of these molecules and to promote peptide exchange. Inspired by these results, MD simulations could illuminate the molecular mechanism of dipeptide binding in chapter 6.

- In chapter 7 free energy perturbation simulations are combined with a generalized Born (GB) model, to rapidly estimate free energy changes associated with alchemical transformations. Molecules are converted by gradually scaling Lennard-Jones parameters electrostatic interactions and Born radii in a series of independent simulations that is coupled by the Hamiltonian replica exchange scheme. While the GB implicit solvent model guarantees for rapid sampling, the replica exchange scheme allows for rapid convergence of free energy estimates.

- However, the chemical structure of biomolecules cannot change in MD simulations. Chapter 8 deals with the trial to mimic a dynamic self (re-)arrangement of chemical fragments during MD simulations. Surrounded by explicit water, these fragments can combine into realistic molecules during a simulation even in presence of a molecular target. High affinity ligands (re-)arrange dynamically within the binding site to ideally form a suitable binding partner. Limits and advantages of this approach will be discussed.

- The final summary includes a recapitulation of the findings of this thesis and a closing outlook.

[†]including the replica exchange method introduced in chapter 3

Chapter 2

Fundamental Concepts

In the following, the fundamental concepts, on which this thesis develops, will be presented. Starting with a short introduction to proteins, this chapter outlines the basics of classical molecular dynamics simulations. Serving as a basis of the replica exchange method in chapter 4, the essentials of elastic network models will be presented. Additionally, this chapter gives an overview on replica exchange methods and free energy calculation techniques that are used throughout this thesis.

2.1 Proteins

Proteins are composed of one or several amino acid chains of varying length. These chains can integrate up to 20 different types of amino acid residues, which share the same scaffold but differ in their side chains (see Fig. 2.1(A)). Encoded on genes in the cell nucleus, information on proteins is transcribed into messenger RNA for transport into cytoplasm. In cytoplasmic ribosomes amino acids polymerize into polypeptides by forming peptide bonds with elimination of water. Atoms of peptide units O, C, N, H

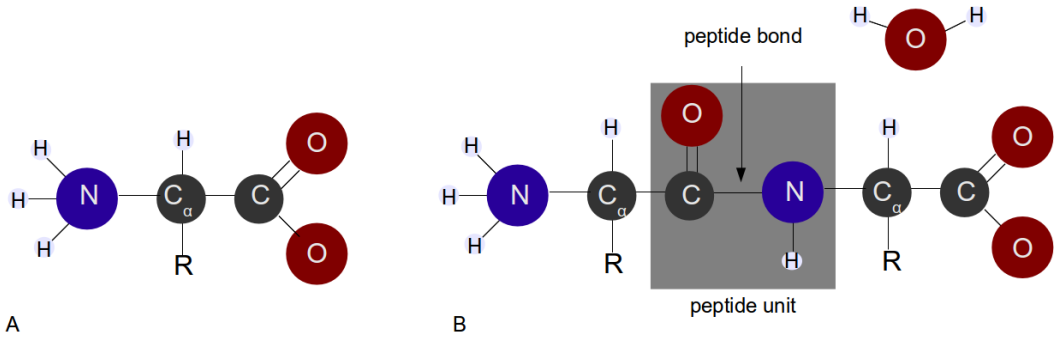


Figure 2.1: (A)Structure of amino acids: An α carbon is bound to a deprotonated hydroxyl and protonated amino group and a rest R, which characterizes the amino acid type. Protonation state accounts for physiological conditions.
(B) Amino acids polymerize to a dipeptide. The peptide unit O, C, N, H forms the planar building block of proteins.

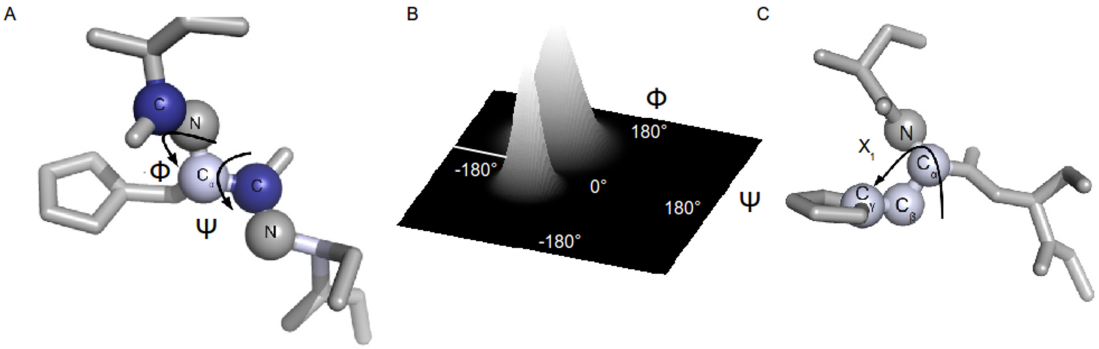


Figure 2.2: (A) Φ-Ψ dihedral angles along the polymer backbone.
(B)Φ-Ψ combinations accumulate in density peaks of the Ramachandran plot.
(C) Side chain dihedral angle χ_1 occurs in 3 rotameric states at $-60°, 60°, \pm 180°$.
Figure adapted from [152]

are fixed within a plane under physiological conditions (see Fig. 2.1(B)). Initially unfolded, proteins find their structure in a multi step process, which can be assisted by chaperone proteins. It is the amino acid sequence–the primary structure–that governs the shape of proteins [4]. Both the type of an amino acid and the characteristics of peptide bonds give structure to a protein: Amino acid side chains differ by charge, spatial extension and degree of hydrophobicity and can serve as hydrogen bond donor or acceptor. While hydrophobic amino acid patterns often form the protein core or peptide binding cavities, hydrophilic amino acids tend to prefer the proximity to water at the protein surface [28].
The folding process is considered as search for the global free minimum in a rough energetic landscape: Proteins never found their functional form within the lifetime of a cell if they would fold in a random process [237]. In contrast, amino acid sequences are supposed to fold into their functional shape adopting intermediate states along a folding funnel. Since each intermediate is characterized by an ensemble of states, there are several folding pathways to a functional form for each amino acid sequence. The shape of proteins is resulting from the interplay of hydrophobic interactions, hydrogen bond formation, electrostatic and Lennard-Jones interactions and conformational entropy. From a structural point of view, proteins consist of secondary structure elements as α-helices [156] and β-sheets [5] that combine with unordered elements to the tertiary structure, which may correspond to their functional shape or may serve as a domain of functional quaternary protein structure. Protein folding or conformational changes are usually associated with both dihedral angle changes along the polymer backbone and polymer side chain flips (see Fig. 2.2(A+C)). As shown in Fig. 2.2(B), polymer backbone and side chains dihedral angle arrange in special patterns: Ramachandran showed adjacent dihedral angles along the polymer backbone to be coupled: Φ-Ψ combinations accumulate in mainly two density peaks of the Ramachandran plot [166], which correspond to characteristic polymer shapes such as α helices and β-sheets[†]. Transitions between these two secondary structure elements include peptide plane flip-

[†]Distinct secondary structure elements form only if at least four (3.7) adjacent combinations are located in the same density peak of the Ramachandran plot. Otherwise an unstructured loop region will develop.

ping: Peptide C, O, N, H units are fixed in a plane under physiological conditions, enabling rotations around $N - C_\alpha$ (Φ angle) and $C - C_\alpha$ (Ψ angle) axes. Moreover, side chain flips contribute to structure formation: Assigned to three distinct side chain angles at $-60°, 60°, \pm 180°$, a change of rotameric states of side chains can result in salt bridge formation between amino acids. Both peptide plane and side chain flips are essential for proteins to adopt their functional form. In chapter 3 and 5 these intrinsic properties are exploited to promote peptide folding and loop refinement.

2.2 Classical molecular dynamics simulations

Analyzing the structure, plasticity and dynamics of proteins in various possible conformational states is of fundamental importance to better understand protein function and its role in biological processes. Experimental methods such as X-ray crystallography or NMR (nuclear magnetic resonance) spectroscopy are widely used to determine the structure of folded proteins at atomic resolution [119]. Even proteins in different stable conformational states can be analyzed, provided the latter can be crystallized or are sufficiently homogeneous to be accessible for NMR spectroscopic studies. Structural changes down to the nanosecond regime can be investigated with fast spectroscopic methods such as FRET [177]. However, the high time resolution is at most limited to the detection of changes in single distances between donor and acceptor chromophores. In contrast to most experiments, MD simulations allow studying the dynamics of proteins simultaneously with high spatial and time resolution [92]. Thus, not only the dynamics of crystallizable proteins or protein conformations but also the dynamics of unfolded and disordered proteins becomes accessible on an atomistic scale. While first calculations on proteins [128] in 1977 gave insights into the dynamics of the (58 amino acids) bovine pancreatic trypsin inhibitor in the picosecond range, folding kinetics of proteins (10-80 amino acids including 3000-11000 water molecules) could be extracted from $\mu s - ms$ simulations in 2011 [114]. Dynamics of proteins is calculated based on Newton's second law:

$$F_i = m_i a_i = -\frac{dV}{dr_i} \tag{2.1}$$

Integrating Newton's equations of motion by means of the Verlet [216] or Leap-frog [67] algorithm yields both coordinates and velocities for each atom per time step. Simulation time scale is limited by the fastest motion in proteins, i.e. vibrating hydrogen bonds that require a time step of 1 fs. The latter can be extended to 2 fs by constraining algorithms like LINCS [64] and SHAKE [101] that readjust bond length between covalently bound atoms. Proteins are represented as a set of atoms with masses m_i, whose geometry and mutual interactions are controlled by the forcefield V, which depends on

a wide range of forcefield parameters:

$$\begin{aligned} V = & \sum_{bonds} \frac{1}{2} k_{ij}^{b} (r_{ij} - b_{ij}^{0})^2 + \\ & \sum_{angles} \frac{1}{2} k_{ijk}^{\theta} (\theta_{ij} - \theta_{ij}^{0})^2 + \\ & \sum_{dihedrals} \sum_{\nu} \frac{1}{2} k_{ijkl}^{\phi} (1 + cos(\nu \phi_{ij} - \phi_{ij}^{0})) + \\ & \sum_{i<j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_{i<j} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \end{aligned} \tag{2.2}$$

with charges q_i force constants $k_{ij}^{b} k_{ijkl}^{\psi} k_{ijk}^{\theta}$, Lennard-Jones parameters A_{ij} B_{ij}, equilibrium bond angle and dihedral values θ_{ij}^{0} ϕ_{ij}^{0} θ_{ij}^{0}. While r_{ij} denotes the distance between atoms i, j b_{ij} corresponds to their bond length. As usually adopting different values, dihedral angles are described by a cosine potential with several minima controlled by the frequency ν. Potentials, which act on not covalently bound atoms, are decisive for correct protein formation: A Coulomb potential includes electrostatic interactions of charged atoms that are approximated as point charges. Interactions between neutral atoms are enclosed by a Lennard-Jones potential, which considers attractive van der Waals attraction as well as Pauli repulsions. In contrast to more accurate quantum mechanical calculations, geometry of proteins does not change: Covalent bonds between atoms are kept fixed during an MD simulation (In chapter 8 we try to break this paradigm by short range potentials that enable dynamic self rearrangement of molecules to change their chemical structure). Forcefield parametrization is crucial for a preferably realistic description of the molecule in its environment [161]. Experimental and quantum mechanical studies result in the common forcefields as AMBER [220], CHARMM [20],GROMOS [213] OPLS/AA [85] that account for the accuracy of molecular dynamics calculations [161].

Interactions of biomolecules with solvent are essential for correct structure formation and protein folding. To adopt physiological conditions, a watery environment as well as different types of ions can be integrated into the simulation. Based on a TIP3P [84] water model, water molecules will be modeled on an atomistic scale (except for chapter 7), by accurately calculating interactions at atomic resolution but at high computational costs. In order to save computational resources, atomistic water molecules and ions can be replaced by continuum solvent models [205]. The solvation free energy $\Delta G_{solv} = \Delta G_{el} + \Delta G_{nonpolar}$ is approximated by electrostatic as well as non polar surface contributions that integrate solvent solute interactions. While $G_{nonpolar}$ reflects the surface accessible area and ,thus, is accounting for the the hydrophobic effect, electrostatic contributions of atomistic biomolecules in ionic solution can be included by solving the Poisson Boltzmann equation:

$$\nabla[\epsilon \nabla \Phi_{el}(r)] = -4\pi\rho - 4\pi\lambda(r) \sum_{i} z_i c_i exp(-z_i \Phi_{el}(r)/(k_b T)) \tag{2.3}$$

with Φ_{el} as electrostatic potential, ϵ as dielectric constant, ρ as charge of the molecule, λ as Stern function, z_i as ion charge and c_i as bulk density of ions. Alternatively electrostatics can be enclosed using the generalized Born approach: Each atom is represented as a sphere with Radius r_i, charge q_i filled with dielectric medium (with

dielectric constant $\epsilon = 1$) that is interacting with the dielectric solvent ($\epsilon = 80$) as implemented in the AMBER package [26, 220]:

$$G_{el} = -0.5 \sum_{ij} \frac{q_i q_j}{f_{GB}(r_{ij}, R_i, R_j)} (1 - \frac{exp(-\kappa f_{GB}(r_{ij}, R_i, R_j))}{\epsilon}) \tag{2.4}$$

with atomistic Radius R_i, charge q_i, r_{ij} as interatomic distance between atom i and j and κ as salt concentration. f_{GB} is a smooth function [‡] [199](for a more detailed description see chapter 7). Explicit hydrogen bonds between water and solute molecules are are not directly included but represented by a strongly polarized dielectric. Viscosity can be integrated using Langevin dynamics[§], which considers solvent friction and random kicks with water molecules. Thus, implicit solvent models include solvent effects into the calculation at low computational costs compared to explicitly integrated water molecules and ions. Having proven to reproduce results of explicit solvent models in test cases, implicit solvent models can be the method of choice, especially if solvent behavior is not of central interest [70]. Implicit solvent models should be applied with care, especially if single water molecules can play a key role in protein folding or stabilization.

2.2.1 Approximating large scale motions of proteins by elastic network model analysis

Global motions of proteins are accessible in MD simulations usually at high computational costs. However, elastic network models (ENM) have proven as alternative to approximate structural flexibility of proteins rapidly [207]. An ENM analysis predicts global motions of large protein domains independently of forcefield details by assigning lowest frequency normal modes (also termed as soft modes) to biological relevant large-scale motions within proteins. As first explored by Tirion [207], softest modes of a protein remain unaffected if Lennard-Jones and electrostatic interactions between atom pairs below a cutoff distance are replaced by harmonic potentials. This is also true on a more coarse grained level, where amino acids are represented by their C_α atoms [65]. Indeed, the ENM model has shown to capture global motions of a set of 20 proteins, which are crystallized in open or closed states [204]. Calculated soft modes agree particularly well with experimentally observed atomic displacements in case of collective protein motions as, for example, hinge bending. Both protein density and a harmonic spring model were included in Hinsen's normal mode analysis [65]: In this method, proteins are represented by their C_α or heavy atoms, which are connected by harmonic springs. Input structure is assumed to be an equilibrium conformation, which is true for crystal structures. Hinsen's normal mode analysis yields fluctuations of C_α-C_α distances around their references. On this frequency scale softest modes characterize the dynamics of large scale inter-domain motions, focused more on the backbone than on the side chain motions. However, amino acid sequence is indirectly

[‡] $f_{GB} = [r_{ij}^2 + R_i R_j exp(-r_{ij}^2/4R_i R_j)]^{\frac{1}{2}}$

[§] instead of Hamiltonian dynamics

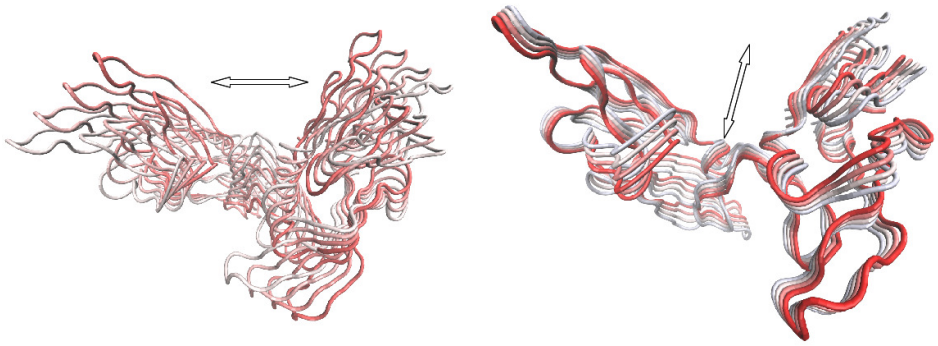


Figure 2.3: ENM modes capture domain motions of proteins.

incorporated into the ENM potential by local C_α atom density. C_α atoms of proteins are interacting pairwise via harmonic potentials, which decrease with $C_{\alpha i}$ -$C_{\alpha j}$ distance.

$$V(R_1, ..., R_N) = \sum_{C_{\alpha i}, C_{\alpha j}} V_{ij}(R_i - R_j) \tag{2.5}$$

with $V_{ij} = k(R^0_{ij})(|r| - R^0_{ij})^2$. A distance dependent force constant $k = \exp(-(\frac{r}{r_0})^2)$ weights interactions of close atoms that are decreasing with their equilibrium distance [66]. For a protein consisting of a number of N C_α atoms, the second derivative of potential energy with respect to the position at equilibrium results in a $3N \times 3N$ Hessian matrix H, whose elements in their mass weighted form $H^* = M^{-1} \times H \times M$ are consistent with:

$$h_{ij} = \frac{\partial^2 V}{\sqrt{m_i m_j}\partial r_i \partial r_j} \tag{2.6}$$

While the eigenvectors of H^* correspond to normal modes their squared eigenvalues yield vibrational frequencies λ_i of atoms. Thus, atoms fluctuate in the direction of calculated eigenvectors with their vibrational frequencies $\nu_i = \frac{\sqrt{\lambda_i}}{2\pi}$.
In chapter 4, ENM is integrated into the replica exchange molecular dynamics framework, to promote large scale motions of proteins in MD simulations. ENM proves to effectively predict conformational transitions of proteins if level of coarse graining raises: two domain proteins are represented by four centroids only (see Fig. 2.3).

2.3 Advanced sampling approaches to study protein dynamics and plasticity

The application of MD simulations for studying biomolecules is limited by the simulation time scale and the accuracy of current force fields. Thus, we will focus on how to improve the sampling of relevant conformational states during MD simulations. Among available advanced sampling methods the replica exchange or parallel tempering methods are the most popular and most widely used techniques.
Peptides and proteins can adopt numerous locally stable conformations possibly separated by large energy barriers. Conformational transitions between stable states can therefore be rare events even on the time scale of hundreds of nanoseconds to microseconds which can currently be covered by peptide simulations [36, 114, 160, 42]. Various

methods have been proposed to overcome the conformational sampling problem during molecular simulations (reviewed in [87]): Energy barriers can be crossed effectively with simulated annealing techniques that open a large conformational space at high simulation temperatures and select low energy states by cooling the system down [95, 215, 21]. However, these high initial temperatures used in simulated annealing approaches may interfere with the presence of explicit water molecules during MD simulations and depending on the cooling process still result in kinetically trapped conformational states. Alternatively, in potential scaling methods the original potential is scaled down or replaced by a soft core potential in order to lower barriers during energy minimization or a molecular dynamics simulation [100, 76, 206, 169, 170]. Conformational flooding [107] and meta-dynamics methods [106] have been designed to specifically enhance sampling along one selected collective coordinate or a set of collective degrees of freedom of a molecular system. During meta-dynamics simulations Gaussian-type potentials along the collective coordinate are added to the energy function continuously to smooth the landscape and ultimately provide a flat energy surface (along the collective coordinate). In the locally enhanced sampling method multiple conformational copies of a selected region of a molecule are generated and their mean field helps overcoming barriers during the simulation [191]. In recent years the accelerated MD methodology has become popular: the force field function is flattened if the potential energy falls below a preset level [61]. This, in turn, reduces relative barrier heights. These approaches accelerate sampling by modifying the temperature or the force field (Hamiltonian), without increasing the computational demand dramatically compared to continuous MD simulations. But modifying the temperature or the Hamiltonian also results in a change of the sampled conformations. This is a drawback especially in a high dimensional coordinate space, where states of little relevance for the temperature or Hamiltonian of interest may be oversampled. Additionally, the free energy difference between folded and unfolded states of a peptide or protein may be shifted. For some of the above techniques it is in principle possible to recover the state distribution at the original energy function. However, this depends strongly on the overlap of the sampling of the deformed potential energy landscape or increased temperature with the original force field or desired temperature. replica exchange simulations in principle avoid this disadvantage, however, at the cost of a significantly (sometimes dramatic) increase in computational demand.

2.3.1 Parallel tempering or temperature replica exchange simulations

The parallel tempering or replica exchange molecular dynamics (REMD) method is one of the most successful and most widely applied methods to enhance conformational sampling in Monte Carlo (MC) [203, 144, 145, 87, 164] and MD simulations (first application in [201]). The technique was originally introduced in simulations of spin glass systems [203] and later extended to simulations of peptides in MC [144] and MD approaches [201]. In REMD simulations several copies (replicas) of the system are simulated independently and simultaneously by means of classical MD or MC methods at different simulation temperatures (or force fields: Hamiltonians, see below). At preset intervals pairs of replicas (usually neighbors in temperature or force

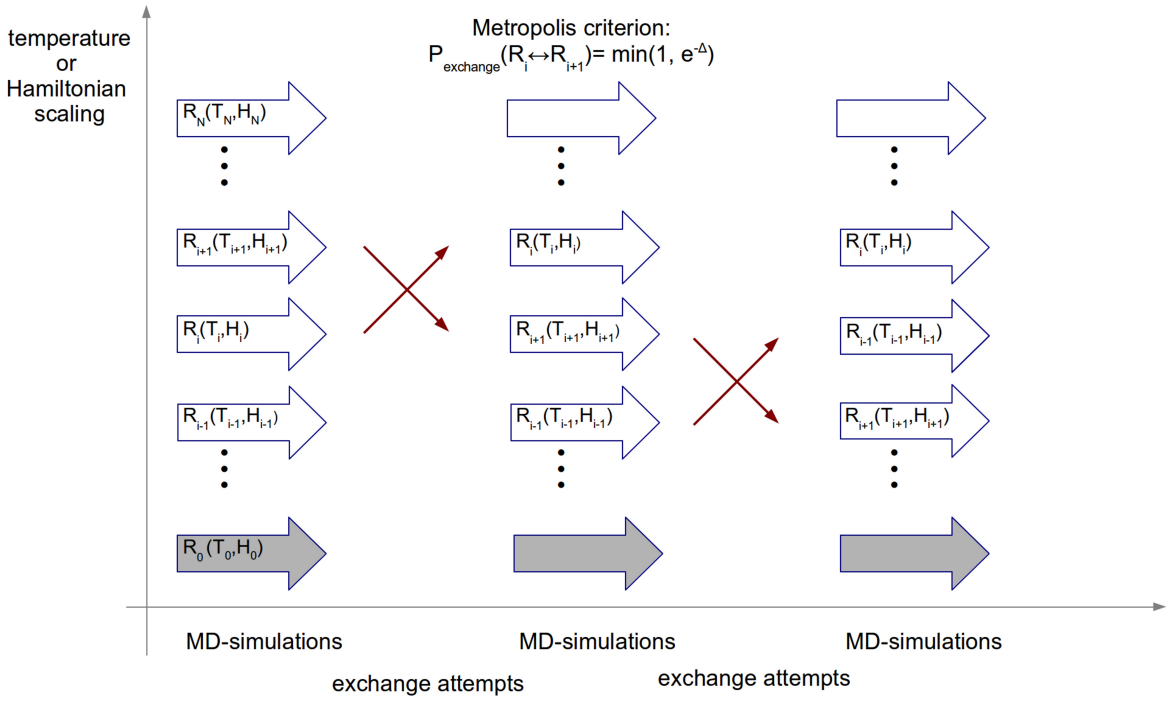


Figure 2.4: Schematic representation of a replica exchange molecular dynamics (REMD) simulation. Each horizontal arrow represents a short simulation interval in between exchange attempts. The temperature or Hamiltonian can vary between the parallel runs (in the vertical direction). The crossed arrows indicate a successful attempt to exchange between neighboring simulations according to an appropriate Metropolis criterion [145].

field modification) are exchanged with a specified (Metropolis) transition probability (schematically shown in Fig. 2.4). In its original implementation temperature is used as a condition to be varied and exchanged among the replicas.

$$\begin{aligned} &w(x_i \to x_j) = 1 \text{ for } \Delta \leq 0 \\ &w(x_i \to x_j) = \exp(-\Delta) \text{ for } \Delta > 0 \\ &\text{where } \Delta = (\beta_i - \beta_j)[E(r_j) - E(r_i)] \end{aligned} \tag{2.7}$$

The random walk in temperature allows conformations that are trapped in locally stable states (at a low simulation temperature) to escape by exchanging with replicas at higher simulation temperature. These exchanges between the replica runs can improve sampling and do not disturb the canonical distribution of sampled states in each replica. Whether REMD indeed improves the sampling of relevant states, depends on the height and type of barriers of the molecular system [236]. If the energy barriers are not higher than the thermal energy per degree of freedom it is unlikely that T-REMD will improve sampling compared to the same number of classical MD simulations running at the reference temperature [236]. The performance of replica exchange Monte Carlo has been directly compared in two different potential landscapes. While in a double well potential parallel tempering can speed up the equilibration time exponentially, only little enhancement is observed in a “golf course” potential [123]. T-REMD will fail if barriers are mostly entropic [123] and sampling can even get worse compared

to regular classical MD for particular systems [171]. The number of accepted exchanges between neighboring replicas is of critical importance during REMD simulations: As different replicas only benefit from each other if a frequent exchange is guaranteed, each replica conformation should visit every replica condition as often as possible. This is an essential prerequisite to enable effective diffusion of replica conformations in temperature space. Acceptance of exchanges between replicas requires sufficient overlap of sampled potential energies. The sampled energy distribution of a given system varies with temperature. As the variance of the energy distribution relative to the mean decreases with the square root of the number of degrees of freedom (becomes sharper with increasing system size), a desired temperature range is only covered if the number of replicas increases with system size. The increased number of replicas requires in turn longer total simulation times to allow effective diffusion of all replicas in the temperature space again increasing the computational demand.
Many efforts in recent years have been directed towards optimizing the temperature distribution among the replica runs in order to ensure efficient diffusion of conformations along the replicas [137, 58, 1]. One standard procedure is the exponential distribution of temperatures: Starting at an appropriate value the temperature difference between different replicas increases with increasing temperature. However, other schemes that involve the heat capacity of the system [211] or optimize the exchange rates on the fly, have also been described [138]. Typically, exchanges are only attempted between neighboring replicas in the ladder. Acceleration of sampling may also be achieved by giving up on this restriction and allowing exchanges between all replicas [24, 19]. The total number of exchanges within a fixed total simulation time can simply be increased by attempting more exchanges per time interval. This is expected to enhance the diffusion in the temperature ladder and in turn may improve the sampling (per simulation time interval). It has also been argued that a minimum time interval of $\sim$ 1ps should be respected to allow sufficient relaxation of the system after an exchange [158]. However, recent studies suggest that increasing the exchange attempt rate may indeed improve the sampling of relevant states [192, 193, 22]. It has also been emphasized that the temperature control algorithm itself (the thermostat) of each replica run can influence the efficiency and accuracy of T-REMD simulations and can create simulation artifacts [175, 174]. Use of velocity scaling schemes, in particular, may not guarantee canonical sampling in each replica [22, 175]. The schemes of performing exchanges between neighbors can also affect the T-REMD performance [116]. Recently, approaches have been developed that simulate a T-REMD scheme without actually performing the T-REMD simulation itself [235, 234]. Such methods can be very valuable to systematically investigate the performance of different REMD protocols.

2.3.2 Variants of T-REMD

Many recent efforts have been devoted to reduce the number of replicas during T-REMD simulations without loss of sampling quality. To avoid an excessive increase in the number of replicas for large peptide or proteins systems, solvent degrees of freedom can be eliminated by using an implicit solvent description (e.g. generalized Born (GB) model) [150] or a coarse-grained description of the protein [29]. However, it is not clear whether the accuracy of current implicit solvent models is sufficient for a realistic description of the structure and dynamics of peptides and proteins [143]. In hybrid

explicit/implicit solvent models the simulation of each replica is performed in explicit solvent that is replaced by a continuum during the exchange [227]. Since the continuum model induces an "average" response of the solvent to the solute conformation, the energy in the exchange criterion may differ significantly from the instantaneous potential energy of the complete system in the standard T-REMD simulation. It remains elusive how this may influence the sampling of conformations in each replica.
Other approaches employ separate coupling of solute and solvent to different heat baths (target temperatures) [31] or separate temperatures of different degrees of freedom of the solute (e.g. soft and hard degrees) [102]. Only the solute reference temperatures or the temperature of selected soft degrees of freedom are varied in each replica. Thus, the number of replicas can be reduced if only essential collective modes of the solute (selected by a principle component analysis) are coupled to the heat bath. Conceptually similar is the self-guided Langevin (SGDL) dynamics: Low frequency motions are accelerated (heated) while keeping the high frequency motions at a lower temperature [225]. This approach can also be combined with replica exchange to accelerate the sampling in a reference simulation without a self-guiding contribution [225]. One should keep in mind that this results in non-physical (non-equilibrium) replica runs. These methods reduce the effective system size that will be compared at each attempted replica exchange. However, exchanges are performed between non-physical replica systems (also the reference replica exchanges with non-equilibrium replicas), which may eliminate the advantage of achieving a canonical equilibrium sampling in each replica. Also, the artificial temperature gradient at the solute-solvent interface (or between different degrees of freedom of the system) may cause unpredictable artifacts within each replica simulation. It is possible to resolve this problem and to effectively increase the temperature of parts of a system using the solute-tempering REMD method developed by Berne and coworkers [119]. Since this involves the scaling of parts of the Hamiltonian it will be discussed below in the paragraph on Hamiltonian (H)-REMD.
The temperature of replica runs can also be periodically changed in a controlled fashion [113, 109, 7]. Exchange attempts are synchronized with the periodic temperature changes such that an exchange is only tried when the temperature difference between neighboring replicas is minimal [109]. Non-equilibrium switches between replica conditions have also been used to reduce the number of replicas in T-REMD [7]. Since in these approaches the temperature differences are significantly larger the relaxation of the system towards an equilibrium might be more relevant than in standard T-REMD and may limit the efficiency (frequency of exchange attempts) of the approach.
Another interesting approach to improve the sampling in T-REMD and to limit the number of replicas is the coupling to a reservoir of conformations generated separately (prior to the REMD simulation typically using an implicit solvent model). Conformers selected randomly from the reservoir can exchange with the highest temperature replica of the T-REMD and can enhance the speed of conformational sampling. Both exchanges with a Boltzmann (equilibrated) or non-Boltzmann reservoir have been described [146, 178, 173] indicating improved sampling compared to regular T-REMD.
For certain applications it is not necessary to cover a large range of different temperatures throughout the complete simulation. For example, the structure of a protein can be refined in T-REMD simulations starting with a few replicas at high temperatures followed by a step wise "cooling" of the set of replicas until the lowest temperature replica has reached a preset low target temperature [89]. Compared to the standard

simulated annealing technique (single conformation) it is possible to keep not only one but a set of promising conformations throughout the REMD simulation until a low temperature has been reached.

2.3.3 Hamiltonian replica exchange approaches

Instead of using the simulation temperature as a replica coordinate it is also possible to use the force field or Hamiltonian of the system as a replica coordinate. A linear scaling of the force field along the replicas this is actually equivalent to a scaling of the temperature (or the beta in the Metropolis exchange criterion). However, the force field of a classical simulation is typically a function consisting of several additive terms. Hence the Boltzmann factor in the Metropolis exchange criterion can be split in its product with respect to each energy term. Therefore, the advantage of H-REMD is the possibility to scale parts of the force field among the replicas (still resulting in canonical sampling in each replica). This is not possible for the temperature which is a function of the complete system. In turn scaling just parts of the Hamiltonian decreases the effective system size considered at an attempted exchange and can reduce the number of required replicas [52]. Several variants of H-REMD have been described and evaluated. It is for example possible to scale just the non-bonded interactions of the solute-solute and solute-solvent interactions among the replicas either linearly [52, 3]. or employing a soft-core potential [74]. Similar in the solute-tempering approach [119] the total potential energy of the system is split into parts that involve the solute-solute, solute-solvent and solvent-solvent interactions separately. In the replica runs the solute-solute and solute-solvent interactions are reduced (linearly) by a preset factor. This selected scaling is equivalent to an effective increase of the simulation temperature of the part of the system that is scaled among the replicas. In this way it is possible to influence the simulation temperature of just the solute only or of both the solute and the solute-solvent interactions. Thus a smaller number of replicas covers the desired temperature range. While the method showed promising results on peptide systems it was less efficient than standard T-REMD in case of protein folding [75]. This approach was recently further refined by scaling only the intra-molecular interactions among the replicas [218].
Other variants of the solute tempering method mostly differ in the type of interactions that are "tempered" or scaled among the replicas [118, 134]: While the interactions between two domains of a protein and the interaction of the domains with the solvent are included in the tempered part, the interactions within each domain remain unaffected. This can improve the sampling of domain-domain arrangements avoiding the sampling of states that affect (unfold or refold) the individual protein domains. Replica exchange has also been combined with accelerated MD simulations (see above) where the force field function below a preset energy is set to a constant level, which is considered as replica coordinate [50]. Besides, conformational transitions can be introduced by scaling van der Waals radii of atoms among replica runs [80]. These methods require careful testing on the degree of deforming the potential and the van der Waals radii, that affect hard degrees of freedom of the system: Large energy differences in the scaled potentials may provoke low exchange acceptance between neighboring replicas.
Other H-REMD approaches focus on reducing energy barriers that are specific for a certain type of biomolecule, e.g. proteins or nucleic acids. In the dihedral angle

biasing-potential (BP-)REMD approach [88] a specific biasing potential promotes peptide backbone transitions. The biasing potential reduces energy barriers associated with peptide backbone dihedral transitions (soft degrees of freedom of the system) along the replica ladder. It was derived from simulations on small peptide model compounds. The level of biasing gradually changes along the replicas such that frequent transitions are possible at high levels of biasing and the system can escape from getting trapped in local energy minima. Since exchanges between replicas are independent of the number of solvent molecules the method requires much fewer replicas for efficient sampling compared to standard T-REMD. The approach was applied to fold several peptide and protein systems [88, 89] and to refine modeled proteins[90]. Related approaches specifically promote dihedral transitions by scaling van der Waals interactions between 1-4 neighbors in peptide bonds [135]. A related BP-REMD enhances conformational sampling of nucleic acid backbone states and promotes coupled transitions of backbone dihedral angles by an adaptive potential along the replicas [34]. This effective biasing potential evolves during an equilibration phase and is subsequently used in a REMD simulation. The approach results in improved sampling of nucleic acid backbone states with fewer replicas compared to T-REMD.

In the resolution exchange method each replica is represented by a linear combination of a fully atomistic molecular description and a coarse-grained model. While one end of the replicas represents only the atomistic model, the other end describes the completely coarse-grained structure [121]. As the coarse graining enhances the conformational search in higher replicas, sampling may also improve in the atomistic reference system. The performance, however, depends on the overlap at the CG level with the relevant states in the all-atomistic representation. Another approach that combines a low-resolution view of a protein with an atomistic simulation requires only a small number (5 replicas) of replica runs: Flexible degrees of freedom of a protein molecule. can be calculated using an elastic network model (ENM). The biasing potential is constructed considering the low resolution flexibility that promotes movements of the protein along its softest degrees of freedom (i.e. eigenmodes with small eigenvalue from the ENM calculations). Thus, large scale motions in protein molecules are accelerated [229].

Meta-dynamics employs Gaussian-potentials to adaptively bias MD simulations along preselected coordinates [106] that are not necessarily Cartesian coordinates of the protein or peptide but can also be variables such as number of hydrogen bonds or radius of gyration or number of atomic contacts. The sampling in meta-dynamics can be improved by applying the biasing potentials only in replica runs and including one replica without a biasing potential. It has been shown that this approach can accelerate folding simulations using only a modest number of replicas [159, 33].

Most of the REMD approaches described in the previous sections have been developed to improve the sampling of conformational states of stably folded proteins or to simulate the folding process. However, many proteins of biological importance are intrinsically disordered or contain segments of disordered structure (adopting no single stable folded structure) [45]. This includes many proteins associated with diseases such as diabetes or Parkinson‘s and Alzheimer‘s diseases. Placed in a flat free-energy landscape without a global minimum, proteins do not adopt a well defined tertiary structure and remain often inaccessible for experimental high-resolution techniques. The conformation of disordered proteins often becomes ordered upon binding to target partner structures.

As a proof of principle extensive unrestrained MD simulations on an intrinsically disordered protein have recently been performed [115]. The extensive simulations on a timescale of 200 μs gave results in reasonable agreement with experimental NMR data demonstrating that current modern force fields are useful to study the conformational dynamics of disordered proteins.
Advanced sampling techniques such as REMD simulations could be helpful to characterize the available conformational space of disordered proteins and are increasingly being used to study such systems (reviewed in [51]). Replica exchange MD methods are widely used to enhance the conformational sampling of peptide and protein molecules. Many efforts in recent years have been devoted to optimize the performance of T-REMD simulations mostly with the aim to reduce the number of required replicas to cover a desired temperature range. A distinct advantage of the H-REMD approach is the possibility to modify only parts of the Hamiltonian or force field among the replicas since it is an additive function of several terms. Such splitting of the force field into a constant part and a part that varies among replicas is not possible for the temperature (without giving up on equilibrium canonical sampling) which is a property of the whole system. Most of the recent method developments focus on the adaption of H-REMD methods to enhance the sampling for specific systems or specific variables of biomolecules (such as collective degrees of freedom or sampling of different dihedral states). The H-REMD methodology has also evolved as a powerful tool to improve free energy calculations and is becoming a standard method to speed up convergence in such simulations. REMD methods could become also of increasing importance for the study of the flexibility and plasticity of disordered proteins and peptides and structure formation upon association because such systems are difficult to characterize experimentally at atomic resolution.

2.4 Free energy calculations

The binding free energy is a key measure to quantify both the stability of folded conformations and the binding affinities of protein-ligand complexes. Thus, a lot of efforts aim at designing computational methods to rapidly predict binding free energies. As experimentally accessible measure it is ideally suited to test computational methods. In this thesis free energy calculations contribute to elucidate the functionality of biological systems:

1. Free energy differences distinguish different conformations of biomolecules to finally characterize the biological active state.

2. Binding free energies of protein-ligand complexes measure binding affinities and are of major impact for drug design and for exploring biological processes.

3. Binding free energies distinguish binding affinities of ligands that arise from one another by alchemical transformations. Bound to their molecular target, chemical groups of ligands are annihilated or created in a series of parallel simulations.

Free energy changes can be extracted from classical MD simulations: While computationally very efficient end state methods as MM/PBSA [99] approximate free energy vales by evaluating single (or several) classical MD trajectories only, techniques as

thermodynamic integration (TI) [96], umbrella sampling(US) [209], or free energy perturbation (FEP) [238] provide rigorous free energy values. Requiring a series of parallel simulations that are bridging the gap between final and end state of molecules this involves the generation of several intermediate states that represent the transformation along the reaction coordinate increasing computational demand. For the convergence of the calculated free energy difference the appropriate sampling of relevant conformational states for each of these intermediates is of critical importance. Exchanges between intermediates often prevent conformations from being trapped and thus improve convergence. To exchange sampled states along the reaction coordinate in free energy simulations is already embedded in the H-REMD exchange scheme. Beginning with the work by Woods and coworkers [223] replica exchange methods are now frequently used to allow exchanges between intermediate states. In many cases, convergence of thermodynamic integration simulations [223, 94, 185] and free energy perturbation simulations [94] improved: not only relative free energy differences in alchemical free energy (due to addition or removal of chemical groups) but also the absolute free energy of complete ligand binding [172, 82, 83, 108] can be calculated. Multidimensional replica exchange (e.g. in temperature and Hamiltonian) improved free energy results as well [83, 200]. In this work Umbrella Sampling [209] (not coupled to H-REMD) and MM/PBSA [56] free energy calculations could give insights into the function of MHC class I molecules in chapter 5 and 6. In chapter 7 free energy perturbation was integrated into a Hamiltonian replica exchange scheme to enhance alchemical free energy calculations in implicit solvent.

2.4.1 Umbrella sampling

Conformational changes of biomolecules are usually coupled to their functionality: In principal, conformational space of a molecule could be sampled completely during a standard MD simulation. If trapped in a rough energy landscape, transformations between inactive and active form rarely occur on current simulation time scales. To overcome this sampling problem, conformational changes can be induced along a predefined reaction coordinate. In a set of classical MD simulations conformations are artificially transformed by means of different harmonic biasing potentials in a set of parallel simulations, determining the pathway through conformational space in Umbrella windows [209].¶: $E_b(r) = E_{ub} + w_i(r)$ with E_b as biasing potential, E_{ub} as unbiased potential $w_i = 0.5k(\zeta - \zeta_i^{ref})^2$ and force constant k. Pulling the system from one state into another, the biasing potential restricts the pathway but does not affect its momenta. The force constant should be chosen with care: While on one hand it must be large enough to overcome energetic barriers along the reactions pathway, on the other hand very narrow Umbrella potentials barely overlap. And it is this overlap, that is decisive for the quality and the convergence of free energy calculations in the WHAM [104] analysis (see below). The probability of finding the system between in

¶In order to enlarge convergence, different windows can even be coupled in the H-REMD exchange scheme (not done in chapter 5)

the interval $d\zeta$ around ζ in the unbiased system yields:

$$\begin{aligned} P_u^i(\zeta) &= P_b^i(\zeta) exp[\beta w_i(\zeta)] < exp[\beta w_i(\zeta)] > \\ Q(\zeta) &= \frac{\int \delta[\zeta(r) - \zeta] exp[(-\beta E)] d^N r}{exp[(-\beta E)] d^N r} \end{aligned} \tag{2.8}$$

The unbiased free energy for window i results in $A_i(\zeta) = -(1/\beta) ln(P_b^i(\zeta) - w_i(\zeta) + F_i)$

$$\begin{aligned} exp(-\beta F_i) &= \int P_u(\zeta) exp[-\beta w_i(\zeta)] d\zeta \\ &= \int exp(-\beta [A(\zeta) + w_i(\zeta)] d\zeta \end{aligned} \tag{2.9}$$

This derivation is exact and requires sufficient sampling per window only. The unbiased free energy F_i is estimated by means of weighted histogram analysis method (WHAM [104]) that considers the total unbiased probability distribution $P_u(\zeta)$ as weighted sum over the unbiased probability distribution $P_u^i(\zeta)$ in each window:$P_u(\zeta) = \sum_i^{windows} p_i(\zeta) P_u^i(\zeta)$. WHAM aims to minimize the statistical error σ of the unbiased probability distribution by an appropriate choice of the weights p_i: $\delta\sigma_i(P_u)/\delta p_i = 0$ respecting the normalization condition $\sum p_i = 1$. In order to provide for a continuous unbiased probability distribution–neighboring Umbrella windows must overlap sufficiently–the number of states sampled per window and the strength of the force constant should be be chosen with care. F_i is calculated iteratively until convergence (for review see [93]). In chapter 5 Umbrella sampling proves successful to demonstrate differences between two MHC class I allotypes. As a drawback, Umbrella sampling results depend on predefined reaction coordinates: Thus, the native pathway is eventually not captured during simulations.

2.4.2 Molecular mechanics Poisson-Boltzmann/ generalized Born surface area method (MM/PBSA, MM/GBSA)

Computational effort reduces in end state methods as MM/PBSA, which post process one single two or three classical MD simulations in explicit solvent [‖]: Dispensing with a huge set of parallel simulations (i.e intermediates), this approach can be applied for ligand affinity predictions, drug design [71, 103] and for free energy calculations of two different conformers. Moreover, it can predict binding poses and binding affinity more accurately than scoring functions during molecular docking [72]. While a gas phase term is respecting intramolecular interactions ΔE_{MM}, solvation free energy is calculated based on an implicit solvent model, which respects polar Poisson-Boltzmann (MM/PBSA) or generalized-Born contributions (MM/GBSA) and nonpolar surface contributions (as described in 2.2). Entropic contributions of the solute are estimated as a rigid rotor harmonic oscillator and extracted from a elastic network model or quasi-harmonic analysis. In general, both, conformational stability and binding free energies can be evaluated using the MM/PBSA approach. While in the first case absolute free energies of two different conformations are evaluated, in the second case relative free

[‖]Principally MM/PBSA also works for single minimized snapshots or crystal structures

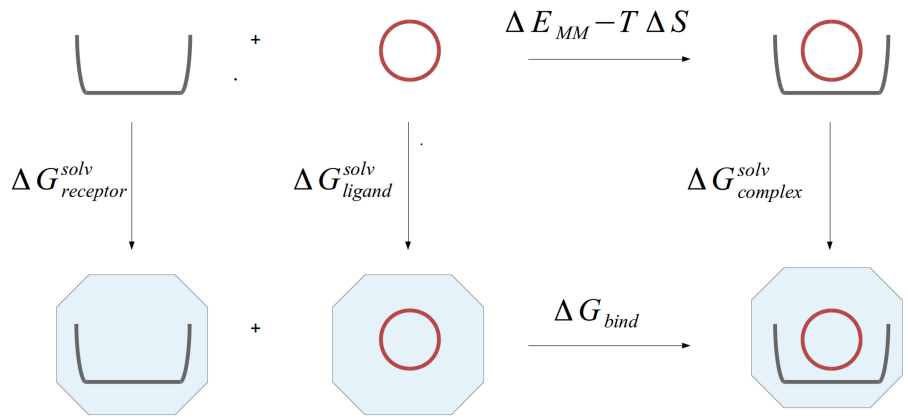


Figure 2.5: MM/PBSA, MM-Thermodynamic cycle for binding free energy prediction

binding energies are evaluated either in a three or a single trajectory approach. The thermodynamic cycle in Fig. 2.5 yields the binding free energy:

$$\Delta G_{bind} = G^{solv}_{complex} - (G^{solv}_{ligand} + G^{solv}_{receptor}) - T\Delta S + \Delta E_{MM}. \tag{2.10}$$

In order to account for maximal accuracy, biomolecules are sampled in explicit solvent. Structures in vacuum are generated simply by removing solvent molecules, resulting in the solvation free energy for the ligand, receptor and complex. If ligand, receptor and complex are simulated in three independent simulations (three trajectory approach), conformational flexibility of receptor and ligand is respected on the classical MD time scale. However, this is not true in a single trajectory approach (i.e. a single classical MD simulation), which restricts the conformational variety of receptor and ligand. As an advantage the angle, dihedral angle and bond terms as well as receptor internal and ligand internal van der Waals and electrostatic interactions cancel out in the force field term ΔE_{MM} and do not contribute to statistical noise in free energy calculations. Due to this fast convergence we apply free energy calculations based on a single trajectory approach in chapter 6.
MM/PBSA MM/GBSA result in "pseudo" free energy values providing an approximation of rigorous free energy values only. Inaccuracies result from the decomposition of ΔG_{bind} in equation 2.4.2, which is based on simplifications as the widely accepted additivity [40] approximation, a rigid rotor approximation and the Sakur Tetrode estimate [136, 11] that is originally derived for ideal non interacting gases in the Maxwell Boltzmann regime. Entropic contributions are roughly estimated by a normal mode analysis (ENM) only, whereby the error grows with flexibility of biomolecules. The quality of entropic estimates additionally depends on broad conformational sampling, which is restricted especially in the one trajectory approach as ligand and receptor are simulated in complex only.[††] Other inaccuracies occur, since the hydrophobic effect is respected by a surface term only. Poisson Boltzmann and generalized Born implicit solvent models often fail predicting polar and buried charges appropriately. Since solvent is regarded as uniform dielectric $\epsilon = 1$, neither its partial charge distribution nor its viscosity or explicit H-bond formation is considered in an implicit solvent approxi-

[††]Sampling and entropy estimates improves in the three trajectory approach described above at the expense of increasing error bars of absolute free energy values.

mation. Moreover, large statistical errors enter into absolute free energy values ΔG_{bind}, which are finally calculated by averaging over the large energy contributions of several snapshots of the MD trajectory: Even small conformational changes of different snapshots can easily lead to high free energy fluctuations [196], provoking large statistical uncertainties and errors that are in order of the calculated energy values. Averages of solvation free energies of the protein, the complex and its parts and the absolute binding free energies of the complex in vacuum (i.e. molecular mechanics term ΔE_{MM} and entropic contributions ΔS) sum up to the final absolute binding free energies in solution. Thus, propagation of uncertainty makes errors grow additionally. In this respect, too, it is more accurate to calculate small ΔG differences directly instead of calculating ΔE and ΔS separately [98].

2.4.3 FREE ENERGY PERTURBATION

In order to provide an effective tool for alchemical transformations as alanine scanning, we aim to combine the speed of implicit solvation with the accuracy of free energy perturbation (FEP introduced by [238]). FEP yields the rigorous Helmholtz free energy difference between two states 1 and 2 of a biological system:

$$\Delta A_{1,2} = -k_B T \ln < \exp -(H_2(x) - H_1(x))/k_b T >_1 \tag{2.11}$$

with Boltzmann constant k_b, temperature T, and structure x. A simulation of a biomolecule with structure x in potential $H_2(x)$ and a reevaluation of structure x in potential $H_1(x)$ is principally sufficient to calculate $\Delta A_{1,2}$. As ΔH is usually fluctuating strongly in FEP calculations, the latter are poorly converging. Hence, these energy fluctuations enforce a fine scaling between initial and end states. A set of simulations can bridge the gap between starting and perturbed state. Molecules are modified in parallel simulations that differ by their Hamiltonians (alchemical potential) only. The fully intact protein or ligand in the lowest (reference replica) is transforming along

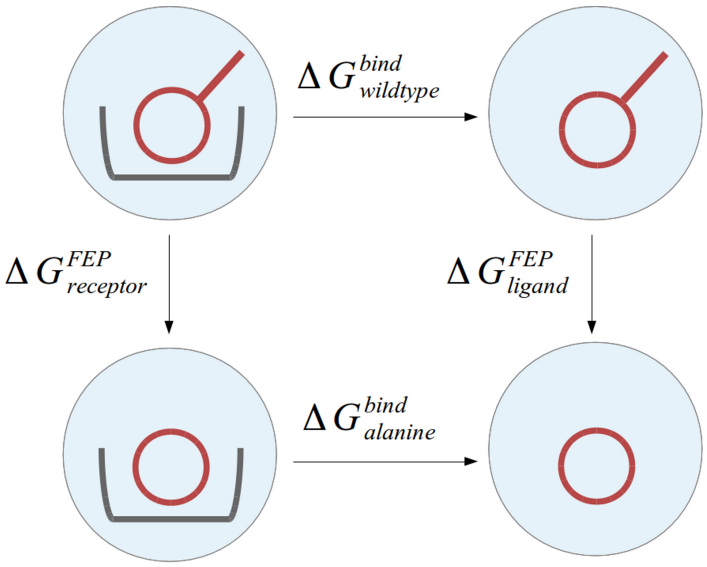


Figure 2.6: Thermodynamic cycle for an alanine scan

the replica ladder: Selected side chains or atom groups are gradually deleted, forming unphysical hybrids to finally mutate into the modified ligand in the highest replica. Summing neighboring free energies results in Helmholtz free energy difference:

$$\Delta A_{12} = -k_B T \sum_i \ln[\exp -(U(\lambda_{i+1}) - U(\lambda_i))/k_b T]_i \tag{2.12}$$

As pointed out in [129], free energy perturbation calculations are already incorporated in the replica exchange framework: In the H-REMD scheme two free energy differences are calculated for exchange partners $H_{i+1}(x_{i+1})$ and $H_i(x_i)$: While energy difference of the conformation x_i in replica H_i and H_{i+1} enters into $\Delta A_{i\rightarrow i+1,up}$, energy difference of the conformation X_{i+1} in replica H_{i+1} and H_i determines $\Delta A_{i+1\rightarrow i,down}$. Forward and backward free energies are a good measure to check convergence of simulations: Ideally $\Delta A_{i\rightarrow i+1,up}$ should equal $\Delta A_{i+1\rightarrow i,down}$. The sum of the free energy differences between neighboring replicas corresponds to the forward and backward transformation free energy of intact to perturbed ligand, respectively.
Relative free energy differences are calculated by exploiting the thermodynamic cycle in Fig. 2.6: The ligand is transformed into its modified structure in its isolated form ΔA as well as in complex with its target protein in two independent series of simulations. The difference of transformation free energy differences yields the relative binding free energy $\Delta\Delta A$.

2.5 Acknowledgements

Parts of this chapter have been published in [152].

Chapter 3

Hamiltonian replica exchange simulations with adaptive biasing of peptide backbone and side chain dihedral angles

Conformational changes of proteins involve dihedral angle shifts. Hence, a biasing potential Hamiltonian replica exchange molecular dynamics (MD) simulation method was designed to specifically enhance these transitions. Two dimensional backbone and one dimensional side chain biasing potentials promote local dihedral angle changes which finally affect the global shape of peptides and proteins and result in improved sampling compared to classical MD simulations. Performance of this methods additionally improves if levels of biasing potentials are adapted appropriately during simulations. Thus, the replica framework is exploited optimally, resulting in both high exchange rates between neighboring replicas and accelerated occupancy/flow of all conformers in each replica.

3.1 Motivation and outline

Classical molecular dynamics (cMD) simulation is a powerful tool to investigate the dynamics of biomolecules [92]. Protein conformations can adopt long lived metastable conformations corresponding to local energy minima in a rough free energy landscape. To characterize the folding process of peptides and small proteins successfully, all atom simulations including explicit solvent have to cover the micro- and even millisecond time range [114]. However, brute-force simulations are often too costly for many applications like structure prediction and structural refinement of protein model structures. Replica Exchange Molecular Dynamics (REMD) methods (introduced by Sugita et al. for MD simulations [201]) have proven successful to speed up conformational sampling (reviewed in [152]). In the standard REMD approach replicas of the system are simulated at slightly different temperatures (T-REMD) and frequent exchanges between neighboring temperature runs are accepted or rejected according to a Metropolis cri-

terion. This can result in enhanced sampling in the replica that runs at the temperature of interest if different replicas exchange frequently and the sampled copies explore all replica conditions. However, the acceptance of exchanges between replicas requires sufficient overlap of sampled potential energies. As the standard deviation of the energy distribution relative to the mean decreases with the square root of the number of degrees of freedom a desired temperature range is only covered if the number of replicas increases with system size. This in turn requires longer total simulation time to allow effective diffusion of all replicas in the temperature space increasing the computational demand. The computational effort can be reduced using the Hamiltonian or force field as replica coordinate [52]: While temperature affects both solvent and solute in the system, scaling of the Hamiltonian along the replicas (Hamiltonian-(H)-REMD) can for example be restricted to the solute (or even special atoms of interest). The resulting modest effective system size during an exchange lowers the required number of replicas. Several variants of the H-REMD approach have been described ranging from scaling non-bonded interactions [52, 3, 119, 118, 135, 80, 79] as well as employing soft core potentials [74]. This included approaches in which only interactions within the solute or solute-solvent interactions are modified along the replicas [3, 119, 118, 135, 80, 79, 74, 75, 134, 218, 223] which raises the effective temperature of the solute only (solute tempering). H- REMD can also be used to improve free energy calculations [223, 94, 120]. The resolution exchange method combines a coarse graining with an atomic resolution model of the peptide along the replica ladder: The fully atomistic reference replica is gradually replaced by a coarse grained representation in higher replicas [121]. Alternatively, it is also possible to embed accelerated dynamics [50] or metadynamics [159, 33] in the replica framework. Global motions can also be enhanced by coupling coarse-grained elastic network analysis (ENM) with atomistic MD simulations in the ENM-REMD approach [229]. In this approach an ENM-based normal mode analysis is performed prior to the REMD simulations to identify collective global degrees of freedom. Based on the ENM-analysis a penalty (biasing or flooding) potential is derived that promotes enhanced global motions along replica simulations [229]. Other approaches focus more specifically on certain types of biomolecules. Conformational transitions of peptides and proteins can be promoted by imposing a suitable dihedral angle penalty potential in each replica on the peptide backbone ϕ and ψ dihedral angles. Inclusion of such dihedral biasing potential along the replicas can enhance sampling in the replica under the control of only the original force field [88, 89]. In the current study this method has been extended by taking advantage of an intrinsic property of proteins: the specific coupling of ϕ and ψ backbone dihedral angles of neighboring amino acids along the peptide backbone. Only limited regimes of the ϕ and ψ dihedral angles in the Ramachandran plot that are separated by energy barriers are actually accessible for the peptide backbone. In order to promote transitions between favorable regimes a 2-dimensional (2-D) biasing potential has been constructed to penalize favorable combinations of ψ and ϕ angles. Different levels of such biasing potentials in the replicas can enhance the sampling of relevant peptide conformations in a reference replica of REMD simulations under the control of the original force field (without a bias). Similar approaches have already been applied successfully to study the backbone fine structure employing BP-REMD simulations with a biasing potential specific for relevant combinations of nucleic acid backbone conformations [34, 91]. The magnitude of the penalty potentials in each

replica is crucial for the performance of the method: While promoting conformational variety high penalty potentials reduce the exchange probability between replicas. Thus, a second development concerns the control of the biasing level in each replica run. In standard REMD simulations the temperature or Hamiltonians of each replica run are set in the beginning of the simulation and are kept constant to achieve appropriate exchange rates between replicas. However, even a high exchange rate between neighboring replicas does not guarantee sufficient traveling of conformers through all replica runs. Frequently, an ordering of conformations in the replica runs results in a loss of performance [152, 217]. To keep a balance between high acceptance rates and broad flux of conformers through the replica runs we designed a protocol that adjusts the biasing potential during the simulation. As the fluctuations of the biasing potentials level off to values below the mean energy per degree of freedom after an appropriate equilibration time, canonical sampling in each replica is not disturbed. Applications of the method to several peptide and protein test cases indicates significantly improved conformational sampling compared to the same number of standard MD simulations at modest computational costs. The initial extended structures of oligo-alanine (Ace-Ala6-Nme), the 10-residue chignolin peptide (sequence: GYDPETGTWG, reference PDB ID: 1UAO [140]) and of the trp-cage protein (reference PDB ID: 1L2Y [69]), were generated using the leap module of the Amber12 package [26]. In case of the crc-SH3 domain of Csk (c-Src specific tyrosine kinase) only the conformation of a loop segment (residues 10 to 20) was randomized by assigning random backbone dihedral angles followed by extensive energy minimization keeping the rest of the structure restrained to the known X-ray structure (PDB ID: 1CSK [14]). For the chignolin, trp-cage and crc-SH3 proteins experimentally determined structure served as reference for comparison with simulation results. Explicit TIP3P water molecules [84] and sodium counter ions were added (oligo-alanine: 2304 waters, chignolin hairpin peptide: 3512 waters, trp-cage :11108, crc-SH3: 4552) to form truncated octahedral boxes. Using the parm03.r1 force field [43], each simulation system was subjected to an energy minimization (5000 steps) with the pmemd module of the Amber12 package [26]. During initial heating and equilibration simulations each peptide was initially harmonically restrained (25 kcal/mol $\AA^{-2}$) to the extended peptide structure. After heating up the system in 100K steps to 300K, the positional restraints were removed gradually. During MD the long range electrostatic interactions were treated with the Particle Mesh Ewald (PME) method [35] and cut off at a real spaces distance of r_{cutoff}=9 $\AA$. The RATTLE algorithm [133] constrains bond vibrations involving hydrogen atoms, which allowed a time step of 2 fs.

3.2 Biasing potentials for peptide main chain ψ and ϕ dihedral angles and side chain χ_1 angles

It is well known that the ϕ (defined by the four atoms of residue l-1 and l: C_{i-1}, N_I, C_{α_I}, C_I) and ψ (formed by: N_I, C_{α_I},C_I,N_{I+1}) main chain dihedral angle pairs in peptides and proteins are coupled and populate specific regimes in the Ramachandran plot. By employing a 2 D penalty potential centered at known stable regimes of the

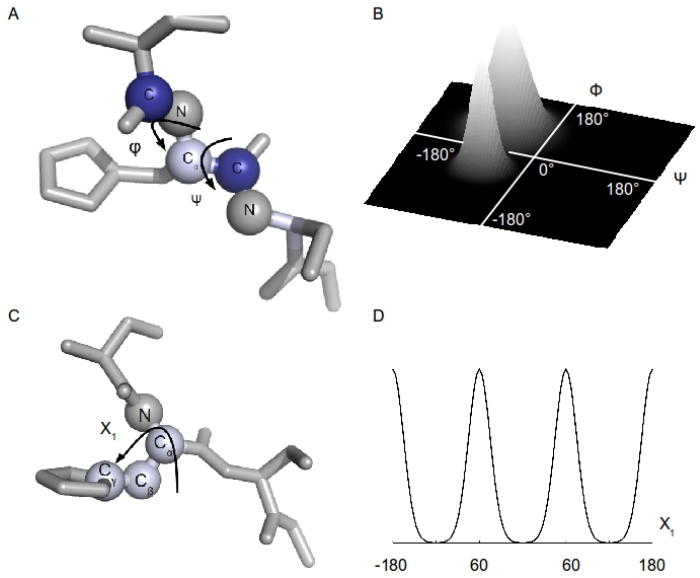


Figure 3.1: (A) Definition of the ϕ-ψ dihedral angles at one residue in a peptide. (B) Shape of the 2D penalty potential that was employed in BP-REMD simulations as replica parameter to destabilize favorable states in the ϕ-ψ dihedral angle (Ramachandran) plot (C) Biasing the χ_1 side chain dihedral angle with a 1D potential (D) destabilizes favorable rotameric side chain states and promotes transitions in replicas.

Ramachandran plot (illustrated in Fig. 3.1B) it is possible to destabilize favorable ϕ and ψ combinations and hence accelerate transitions over energy barriers to alternative peptide backbone substates. As depicted in Fig. 3.1A α-helical and β-sheet density peaks were approximated by a symmetric penalty function with a circular plateau and diameter 2r, where it takes its maximal energy value. Centered around (ϕ_c, ψ_c) the penalty potential decreases continuously on its edges and equals zero at an angular distance of R.

$$\begin{aligned} x &= \sqrt{(\phi-\phi_c)^2+(\psi-\psi_c)^2} \\ V(\phi,\psi) &= E_{pen}^{max} \text{ if } x < r \\ V(\phi,\psi) &= \frac{E_{pen}^{max}}{(R-r)^4}[(x-r)^2-(R-r)^2] \text{ if } r < x < R \\ V(\phi,\psi) &= 0 \text{ if } R < x \end{aligned} \tag{3.1}$$

Test simulations indicated that the β-sheet regime was optimally destabilized by a 2D biasing potential centered at ϕ_c=95°, ψc=120° and with 2r=90° and 2R=130°. In case of the narrower α-helical regime the following parameters were used: ϕ_c=60°, ψc=145o and 2r=40° and 2R=60°. Stable peptide and protein substates may not only depend on backbone dihedral states but also on the conformation of side chains. In order to promote transitions between different rotameric states of side chains, we additionally imposed a 1D penalty potential, that penalizes χ_1 angles (except Pro, Gly, Ala) in the stable-gauche (-60°), gauche (60°) and trans (180°) regimes (illustrated in Fig. 3.1C+D). For simplicity the same penalty for each stable χ_1 state was used (with 2r=40° and 2R=80°). The resulting total penalty potential is a sum of a 2D-term

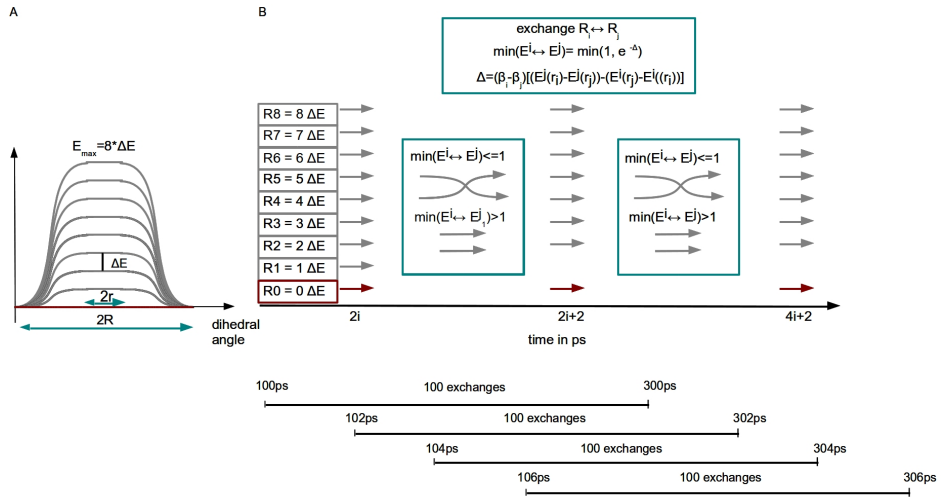


Figure 3.2: Schematic representation of the BP-REMD simulations with different levels of biasing potentials (shown in Fig. 3.1) illustrated in (A) that followed the protocol in (B). Exchange trials were recorded via running windows (100 and 500 exchange trials respectively).The average occupancies as well as exchange rates served as basis for evaluating the biasing levels (ΔE) in the replica runs (see Methods for details).

acting on the peptide backbone and a 1D potential that destabilizes the χ_1 dihedral angle of side chains.

3.3 BP-REMD SIMULATIONS AND ADAPTIVE BIASING

A BP-REMD simulation consisted of a set of 9 parallel replica MD simulations with different levels of the biasing potential described above (of the form illustrated in Fig. 3.2). One of the replicas in a BP-REMD simulation ran with the original force field without biasing potential and serves as the reference replica for subsequent analysis. Exchanges between neighboring replicas were attempted every 1000 steps (2 ps) and accepted or rejected according to a Metropolis criterion (illustrated in Fig. 3.2). Since conformational sampling improves only, if each replica simulation visits each penalty potential frequently, the number of accepted exchanges is crucial for the performance of this method. While large biasing penalties drive conformational transitions effectively, exchanges rates may decrease drastically due to high penalty potential differences between replicas that enter into the exchange criterion. The height of the dihedral angle potential was therefore controlled and appropriately adjusted during the REMD simulation respecting the following criteria: 1. (acceptance criterion) Exchange rates between neighboring replicas shall surpass 20 %. 2. (occupancy/flow criterion) Every starting conformation should visit every replica condition frequently. Since exchanges can occur mainly just between two specific neighbors these criteria are not equivalent. To determine the energy potential on the fly, two time windows are introduced the biasing potential energy differences between the different replica conditions are adapted according to the following criteria: Averaging over the last 100 exchanges, 8 exchange rates between neighboring replicas are evaluated. While biasing potential differences were lowered by 0.5%, if one exchange rate fell below 20%, the biasing increased (by

0.5%), if all exchange rates surpassed 40%. The second occupancy criterion was controlled by labeling each replica start conformation and monitoring the frequency of occurrence of these labeled structures in each replica condition. If the population of each labeled configuration in each replica fell below 2% within the last 500 exchanges the biasing levels were further decreased by 0.5%. Note, that at each scaling step of the biasing potentials the same scaling was applied to each biasing potential in the replica runs. In all test cases the biasing potentials reached average levels after ∼2ns and the fluctuations of replica biasing potentials per dihedral angle pair fell below the mean energy per degree of freedom in all cases (RT: 0.59 kcal/mol, table 3.1) without any residual drift.

replica number	Oligo-Ala E_{pen}	St.Dev.	Chignolin E_{pen}	St.Dev.	Trp-cage E_{pen}	St.Dev.	SH3-loop E_{pen}	St.Dev.
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.2	0.01	0.1	0.01	0.1	0.01	0.1	0.02
3	0.4	0.03	0.3	0.03	0.2	0.03	0.2	0.05
4	0.6	0.04	0.4	0.04	0.3	0.04	0.3	0.07
5	0.8	0.06	0.6	0.06	0.4	0.06	0.5	0.10
6	1.0	0.07	0.7	0.07	0.5	0.07	0.6	0.12
7	1.2	0.09	0.8	0.09	0.6	0.09	0.7	0.15
8	1.4	0.10	1.0	0.10	0.7	0.10	0.8	0.17
9	1.6	0.12	1.1	0.12	0.8	0.11	0.9	0.20

Table 3.1: Biasing potential levels in each replica in the final BP-REMD simulation stage. E_pen and St.Dev. correspond to mean biasing potential levels (in each replica) and standard deviations, respectively, for each of the four test systems (in kcal/mol) not including the first 2 ns. The maximum biasing level corresponds to the plateau biasing level for the dihedral angles illustrated in Fig. 3.1.

3.4 Folding of oligo-alanine, chignolin, and Trp-cage and loop refinement of a SH3 domain

The performance of the BP-REMD method with a 2D-backbone biasing potential along the replicas was first tested on an oligo-alanine peptide (Ace-(Ala)6-Nme) and compared to continuous (c)MD simulations. The starting conformation was a fully extended peptide structure. There is experimental evidence that oligo-alanine peptides adopt mainly a polyproline II conformation in solution [186]. However, the parm03.r1 force field [184] has a slight bias in favor of an α-helical conformation for the oligo-alanine peptide. Therefore, an α-helical peptide served as reference structure for comparison with the simulations. The reference replica simulation of a BP-REMD with 9 replicas was compared with 9 independent cMD simulations that differed in initial starting atom velocities (but using the same start structures as in the BP-REMD simulations, this scheme was used also for all other test cases described below). During the BP-REMD simulations the biasing potential levels were controlled and adapted in order to result in both a high acceptance rate for exchanges and a reasonable distribution of sampled states over the replicas. The root-mean-square deviation of the backbone

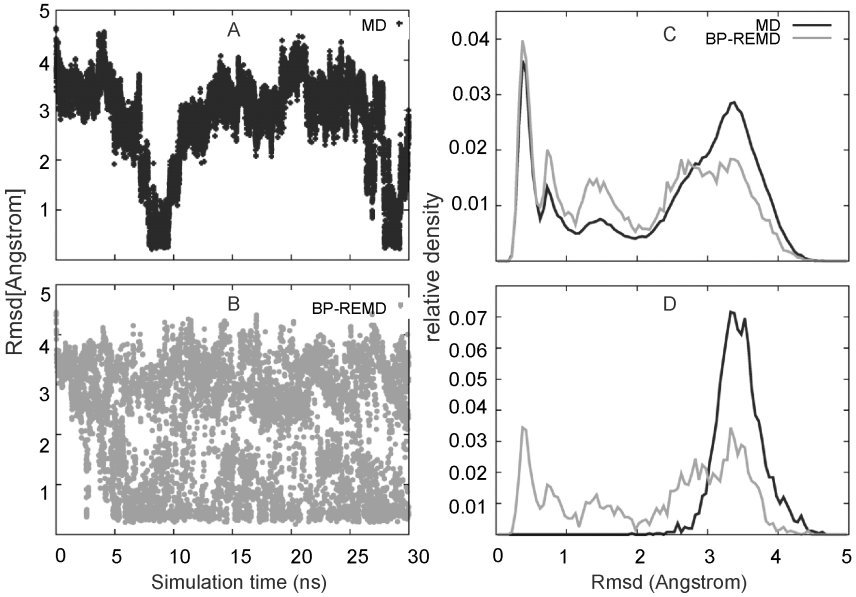


Figure 3.3: The Root-mean-square deviation (Rmsd) of the peptide backbone $\text{Rmsd}_{backbone}$ with respect to an α-helical structure vs simulation time was recorded for the cMD simulations (typical example shown in A) and the reference replica of the BP-REMD simulation (in B). The histogram for the $\text{Rmsd}_{backbone}$ distribution (with respect to an α-helix) is indicated in C, for all 9 combined cMD simulations (black line, entire simulation time) and the reference replica of the BP-REMD simulation (grey line, entire simulation time). The same is shown in (D) for the first 5 ns of all cMD simulations and for the reference replica of the BP-REMD.

($\text{Rmsd}_{backbone}$) with respect to an α-helical reference structure decreased eventually below 1 $\mathring{A}$ in both the cMD (Fig. 3.3A) and the BP-REMD simulation (reference run: see Fig. 3.3B). While an α-helical like structure was already observed after 1 ns in the reference replica of the BP-REMD it appears after 8ns in one out of 9 cMD trajectories. Furthermore, in case of BP-REMD the Rmsd distribution of the sampled states after 5 ns closely resembled the Rmsd distribution after 30 ns in the reference replica (compare Fig. 3.3C and Fig. 3.3D). The collection of all 9 cMD simulations reproduced an equivalent distribution (Fig. 3.3C) indicating that the protocol did not disturb the canonical sampling of states. In contrast, the distribution in all cMD simulations after 5 ns differed significantly from the distribution after 30 ns (Fig. 3.3D). A cluster analysis can help identifying the most likely regime of conformations in a trajectory (the one of lowest free energy) as the most populated cluster of conformations. The most populated and the cluster of lowest $\text{Rmsd}_{backbone}$ from the α-helical peptide conformation were recorded along the trajectory of the reference BP-REMD run (under the control of the original force field). Once the lowest $\text{Rmsd}_{backbone}$ cluster equals the most populated cluster, the native like structure could in principle be identified without experimental reference. For oligo-alanine this was the case after a simulation time of 5 ns data not shown). Moreover, the representative characteristic cluster structure is in good agreement with the reference structure ($\text{Rmsd}_{backbone} < 0.5$ $\mathring{A}$, data not shown).

A more challenging system compared to oligo-alanine is the chignolin peptide (10 amino acids, see Methods) which forms a stable hairpin and has been subject of several

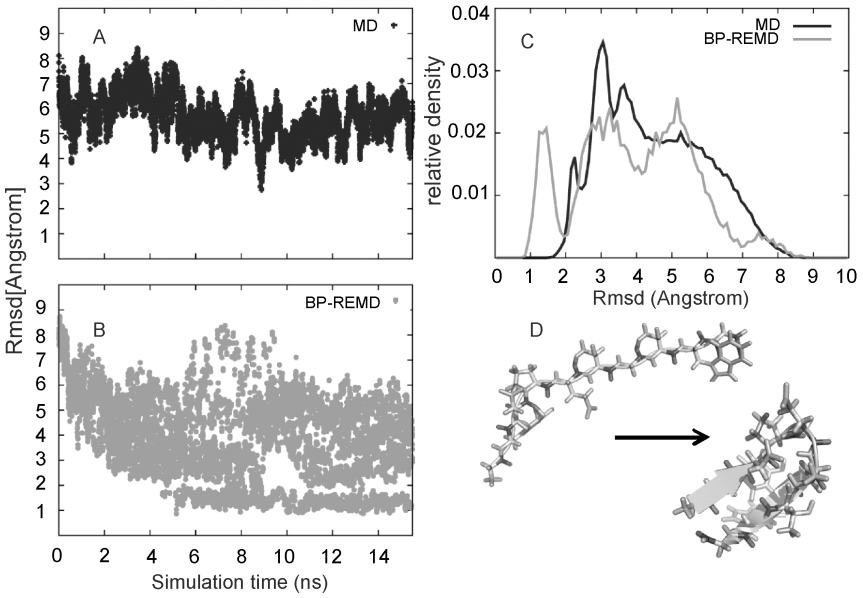


Figure 3.4: (A) $\text{Rmsd}_{backbone}$ of chignolin peptide backbone with respect to the experimental structure vs simulation time was recorded for the cMD simulations (one out of 9 examples) and for the reference replica of the BP-REMD simulation (in B). The histogram for the $\text{Rmsd}_{backbone}$ distribution (with respect to native chignolin) is indicated in C, for all 9 combined cMD simulations (black line, entire simulation time) and the reference replica of the BP-REMD simulation (grey line, entire simulation time). (D) Comparison of a the extended starting structure (left stick model) and native chignolin structure (stick model with beta-hairpin indicated as cartoon).

previous simulation studies in explicit water [92, 88, 186, 184]. In previous cMD simulations it has been possible to fold chignolin starting from a random coil conformation with a folding time of 0.6 μs at a temperature of 340K [92]. In a T-REMD study with 16 replicas random coil starting conformations of chignolin folded after around 150 ns to structures close to the native conformation [184]. In the present study the chingolin peptide was investigated with BP-REMD simulations in explicit solvent performing a similar set of simulations as described for oligo-alanine. However, none of the 9 independent cMD simulations resulted in the sampling of near-native chignolin conformations (best $\text{Rmsd}_{backbone} > 2$ $\AA$, Fig. 3.4A). For the BP-REMD first transitions from fully extended chignolin to conformations close to native ($\text{Rmsd}_{backbone} < 2$ $\AA$, Fig. 3.4B) accumulated after 5ns in the reference replica (see Fig. 3.4C). The peak of the $\text{Rmsd}_{backbone}$ distribution of 1 $\AA$ indicates an accumulation of the folded structure in the reference replica (see Fig. 3.4D). Already after 7 ns of the BP-REMD the most populated conformational cluster represented also the cluster with conformations of lowest $\text{Rmsd}_{backbone}$ relative to the native structure (data not shown). The Trp-cage is one of the smallest (20-amino acid) proteins that contains different secondary structure elements and forms a well defined stable folded structure with a small hydrophobic core [140] (Fig. 3.5) and has already been the subject of several previous simulation studies [92, 89, 165, 86]. While near native structures were sampled after 20 ns in the BP-REMD, no folding to near-native structures was observed in any of the 9 independent control cMD simulations of the same length (each 60 ns, Fig. 3.5A+B).

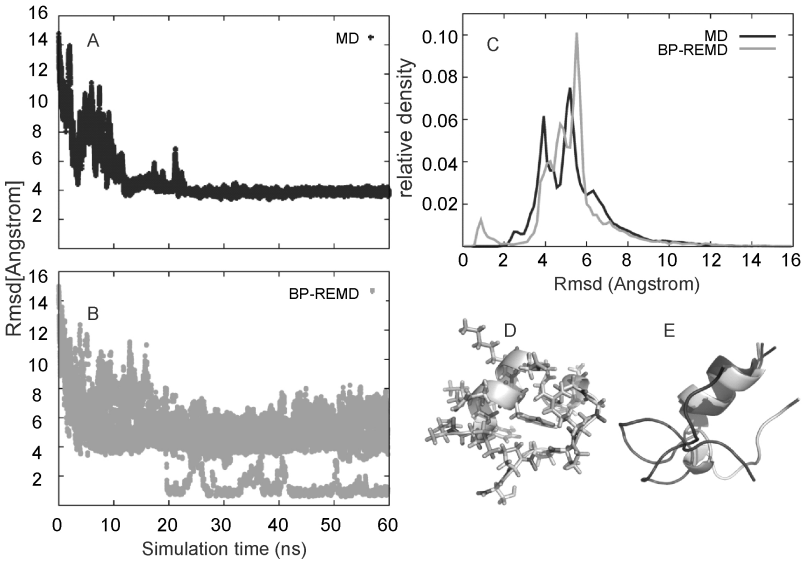


Figure 3.5: (A) The $\text{Rmsd}_{backbone}$ of trp-cage backbone with respect to the experimental structure vs simulation time was recorded for the cMD simulations (one out of 9 examples) and for the reference replica of the BP-REMD simulation (in B). The histogram for the $\text{Rmsd}_{backbone}$ distribution (with respect to native trp-cage) is indicated in C, for all 9 combined cMD simulations (black line, entire simulation time) and the reference replica of the BP-REMD simulation (grey line, entire simulation time). (D) Comparison of the native structure of trp-cage (pdb1Y2T, cartoon and stick model) and a superposition of the near-native backbone conformation (grey cartoon) and several other conformations obtained in the replicas at the final stage of the BP-REMD simulation (darker grey cartoons).

After 41 ns the population of lowest $\text{Rmsd}_{backbone}$ cluster is approaching the one of the most populated cluster (data not shown), however, it is not forming the most populated cluster over the entire trajectory (see Fig. 3.5C). This could be due to insufficient accuracy of the force field. However, to achieve an accumulation of folded structures in quantitative agreement with experiment would require transitions to near- native structures in most of the replicas which requires longer simulation times even with the present BP-REMD technique. Interestingly, at the final stage of the simulation (60 ns) 25% of structures in the replicas contained the native-like α-helix and include in part already native contacts between residues of the helix and the C-terminus of the protein (Fig. 3.5D+E). As a final example the BP-REMD approach was applied to the refinement of a peptide loop structure (illustrated in Fig. 3.6). Predicting the conformation of loop structures is a common problem while modeling protein structures based on homology to a protein with known structure. While frequently part of the structure can be predicted with high confidence other parts, e.g. loops that connect secondary structure elements, cannot be modeled based on homology to a template structure [147]. Most realistic force field models include presence of explicit water and should principally enable very accurate loop structure prediction. A long loop connecting two β-strands in a small SH3 domain protein was chosen as test system. The loop segment (Residues 10 to 20) of the crc-SH3 domain high-resolution X-ray structure :PDB ID:1CSK[84]) was deformed randomly to a $\text{Rmsd}_{backbone} > 6$ $\AA$ relative

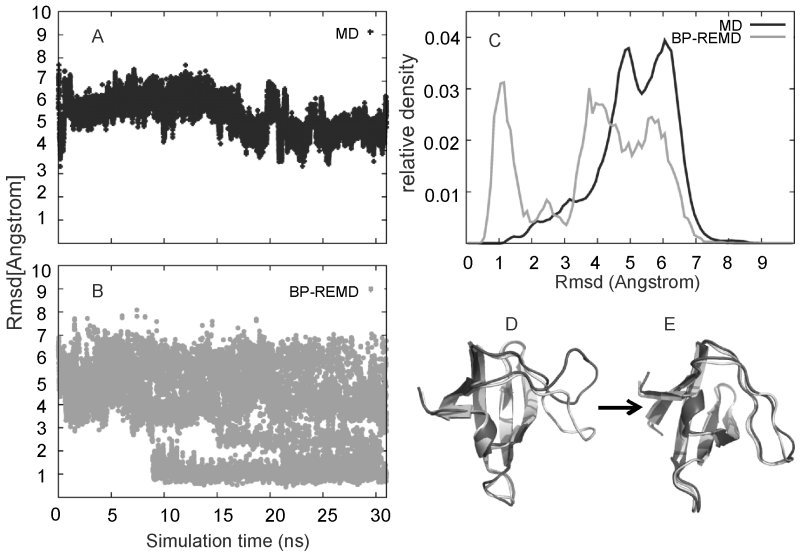


Figure 3.6: (A) $\text{Rmsd}_{backbone}$ of a 12 residue loop segment of an SH3 domain (residue 10 to 20 of PDB ID:1CSK [14]) with respect to the experimental structure vs simulation time was recorded for the cMD simulations (one out of 9 examples) and for the reference replica of the BP-REMD simulation (in B). The $\text{Rmsd}_{backbone}$ of the loop segment was calculated after superposition of all SH3 residues (except for the loop segment) onto the native structure. The histogram for the $\text{Rmsd}_{backbone}$ distribution (with respect to native loop structure) is indicated in C, for all 9 combined cMD simulations (black line) and the reference replica of the BP-REMD simulation (grey line). (D) Comparison of the start structure (black) and native SH3 structure (light grey). (E) final structure (black, representative structure of highest populated cluster) superimposed onto native structure (light grey).

to experiment. During BP-REMD simulations the protein backbone (except for the loop region) was weakly restrained to the native SH3 domain structure (see Methods). Already, after 10 ns conformations in the reference replica approached the native loop structure ($\text{Rmsd}_{backbone} < 1\ \mathring{A}$) and formed the dominant conformational state at the final stage of the (30 ns) simulation (Fig. 3.6B-D). In contrast, no structures with an $\text{Rmsd}_{backbone} < 2\ \mathring{A}$ were observed in standard cMD simulations (9 independent runs of 30 ns, Fig. 3.6A).

The biasing levels in each replica of the BP-REMD simulations were adjusted to guarantee high acceptance rates and simultaneously a broad exchange (occupancy) of each replica conformation in each replica run (see Methods for details). At each adjustment step the biasing potentials of the replicas were uniformly scaled. In every case the automatically adjusted biasing potentials reached a constant level with modest final fluctuation (Fig. 3.1). In the oligo-alanine case a high-acceptance rate and very uniform distribution of conformers in all possible replicas was observed (Fig. 3.7 A). Note, that these two criteria together determine a high traveling rate of the conformers to visit all replica conditions and both were controlled during the BP-REMD simulation (see Methods). For example a high exchange rate between neighboring replicas can still indicate that just two labeled conformers exchange back and force (without visiting other biasing potential conditions). However, this would result in a very non-uniform

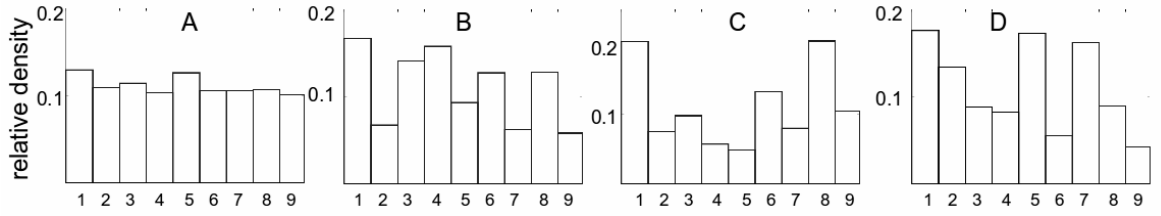


Figure 3.7: Distribution of different starting structures in the reference replica. Each starting structure at the beginning of the BP-REMD simulations was labeled (s1-s9). The frequency of occurrence in the reference replica (bars) reflects its exchange between different replica conditions:
(A) BP-REMD of oligo-alanine, see Fig. 3.3;
(B) BP-REMD of chignolin, see Fig. 3.4;
(C)BP-REMD of trp-cage, see Fig. 3.5;
(D) BP-REMD of 10 residue loop in crcSH3 domain (see also Fig. 3.6).

distribution of the conformers among the replicas. For the other three test cases high exchange rates ($> 40\%$) between neighboring replicas but no completely uniform distribution but nevertheless a full coverage of all (labeled) start structures over all replicas was achieved (Fig. 3.7 B-D).

It should be noted that a time dependent biasing potential creates in principle a non-equilibrium force field condition in each replica (except for the reference replica). The Monte Carlo exchanges between non-equilibrium ensembles can also perturb the sampling in the reference replica. Especially, for a drifting biasing potential (in the initial phase of the BP-REMD) this may result in sampling of a non-canonical distribution in the replicas. However, after an equilibration phase ($\sim$2ns) the biasing potentials reached on average constant levels with 10% fluctuation (table 3.1). Such fluctuations

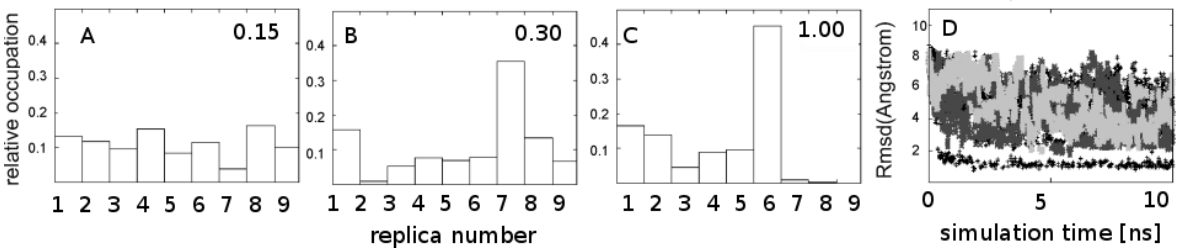


Figure 3.8: Fixed biasing potential levels in each replica: To investigate the influence of the dynamical adjustment of the potential size, BP-REMD simulations were performed with constant dihedral biasing levels in each replica condition (the difference in maximum biasing potential (kcal/mol) per dihedral is indicated). Occupancy of each labeled start structure in the reference replica was recorded in the upper three panels (the change in biasing level between neighboring replicas is indicated). Deviation from desired 11.1% occupancy increases with increasing difference of penalty potentials in neighboring replicas. The lower panel indicates the backbone Rmsd of sampled states in the reference replica for the three different cases (black symbols: ∇E_{pen}= 0.15 kcal mol;dark grey symbols: ∇E_{pen}= 0.3 kcal/mol; light grey symbols: ∇E_{pen}=1.0 kcal/mol.

(similar to temperature fluctuations in a standard MD simulation) may not perturb the average sampling severely because the average Boltzmann probabilities stay constant (if one neglects higher order influences of the fluctuations on the Boltzmann probabilities beyond a linear term). The appropriate choice and adjustment of the biasing potentials is of critical importance for the performance of the method. For the chignolin case this is demonstrated by running BP-REMD simulations with constant biasing potential levels in each replica. In the first BP-REMD the biasing potential levels between neighboring replicas differed by 0.15 kcal/mol which is close to the optimal level found in the final stage of the BP-REMD of chignolin with adaptable biasing levels (table 3.1. As shown in Fig. 3.8A, in this case the starting structures 1 to 9 distribute relatively uniformly in the reference replica indicating that all conformers occasionally visit the reference replica and contribute to improved sampling. As a result successful folding occurred rapidly already after ∼1 ns simulation time (Fig. 3.8D). In contrast, larger differences in biasing levels provoke an ordering of conformations in the replicas preventing all conformations from visiting the different replica conditions frequently and uniformly (Fig. 3.8B+C). As a consequence transitions to near-native chignolin conformations occurred in these cases on longer time scales (∼10-15 ns, not shown).

3.5 Conclusion

Replica exchange MD simulations can improve sampling of conformational states of biomolecules especially in cases where favorable states are separated by significant energy barriers. However, the computational demand of the standard T-REMD approach grows rapidly with system size limiting the method to small proteins or implicit solvent models. Modifying the Hamiltonian (force field) among the replicas instead of the temperature can reduce computational demand. In the current approach the force fields along the replicas differed by both a 2D biasing potential specifically designed to promote backbone dihedral angle transitions of a peptide as well as a 1D biasing potential that affects the first dihedral angle of the peptide side chains. For several test systems the approach significantly accelerated structure formation processes and increased the variety of sampled configurations compared to the same number of independent cMD simulations. Dynamic adjustment of the biasing potentials during simulations constitutes a second advantage over previous BP-REMD approaches with a 1D biasing potential [88]. In this case an adjustment of the biasing levels in each replica on test systems or during short test simulations was necessary [88]. Both high acceptance rates and a broad distribution or flow of conformations among the replicas can be of critical importance for effective sampling: Conformers may tend to "order" during replica exchange simulations such that only a fraction of the replicas is connected through replica exchanges [217]. Dynamic adjustment of the biasing potential in each replica ensures both high exchange rates and good mixing of sampled structures in the replicas. It avoids evaluating the optimal biasing potential levels by test simulations [88] or underestimating its level to ensure high exchanges rates and thus minimizes the risk of reducing the effect of the biasing potential. Meeting the requirements for structural refinement, the BP-REMD method can be applied to selected parts of the protein and thus selectively enhance the conformational sampling of a desired region of a protein. Successful application to a 10 residue loop region in a crc-SH3 domain indicates that

efficient refinement of incorrectly modeled loop structures in explicit solvent is possible at modest computational demands. It is possible to extend the approach to bias additional side chain dihedral angles which could be especially important in case of large bulky residues such as tryptophan. Another extension could be the biasing of the ω-backbone dihedral transition that controls the equilibrium between the backbone cis and trans conformation at proline residues. Such transitions are rarely observed on the time scale of continuous MD simulations although important for the folding of several proteins.

3.6 Acknowledgements

Parts of this chapter have been published in [153].

Chapter 4

Hamiltonian Replica Exchange Combined with Elastic Network Analysis to Enhance Global Domain Motions

The biasing potential method presented in chapter 3 enhances conformational sampling of peptides and proteins by inducing backbone and side chain dihedral angle transitions. These local changes affected the global shape of proteins and promoted folding of test peptides and loop regions. However, the effectivity is coupled the the number dihedral angles that are to be penalized and thus decreases with protein size. As an advantage, flexible regions of the protein can be penalized only which is a benefit for loop refinement. A bunch of biological processes is associated with the motions of rigid protein domains. The dynamics of their flexible hinge regions could be promoted by the biasing potential method. However, the directions of global mobility of rigid protein domains are captured well by coarse-grained elastic network models (ENM) of proteins that are based on a low resolution representation of protein dynamics. The computational effort decreases by means of a Hamiltonian replica exchange molecular dynamics (H-REMD) approach that combines an ENM analysis with atomistic explicit solvent MD simulations: A set of centers, that represents rigid segments (centroids) of a protein is subjected to an ENM analysis. Based on this results, a distance-dependent biasing potential is constructed to promote and guide centroid/domain rearrangements and is integrated into the replica framework. The magnitude and the form of the biasing potentials are adapted during the simulation, basing on the average sampled conformation, to reach a near constant biasing in each replica after equilibration. This allows for canonical sampling of conformational states in each replica.

4.1 Motivation and outline

Classical molecular dynamics (cMD) simulations are increasingly being used to characterize the dynamics of biomolecules [114, 214]. However large-scale motions of

biomolecules often occur only rarely even on the time scale of hundreds of nanoseconds that are currently possible. Parallel tempering or replica exchange molecular dynamics (REMD) simulations is an established technique used to enhance conformational sampling during MD simulations [reviewed in [152]]. Conformational variety is generated in parallel replicas running at different temperatures. At preset intervals (usually neighboring) pairs of replicas exchange with a specified transition probability. Instead of changing the temperature among the replicas it is also possible to modify the force field or selected force field contributions along the replicas (termed Hamiltonian-(H)-REMD). A variety of H-REMD techniques to enhance conformational sampling of biomolecules have been described by scaling different force field terms [52, 3, 118, 121, 88, 159, 80, 34, 91, 153]. In most of these approaches specific force field terms are modified or biased along the replicas such as non-bonded interactions [3, 118, 121, 80] hydrogen bonds and short range contacts[159] or dihedral angle barriers [88, 159, 34, 91, 153]. These approaches can significantly enhance sampling of local changes in proteins but are not specifically aimed at improving the sampling of global domain motions in proteins. In recent years it has been shown that soft normal modes obtained from Elastic network models (ENMs) of proteins frequently overlap with experimentally observed global conformational changes in proteins [207, 6, 65, 204, 208]. In an ENM the mobility of a residue or protein segment depends on the local density and number of short range contacts (usually between C_α atoms of the protein). Collective degrees of freedom can be calculated very rapidly from an ENM model of a protein by a normal mode calculation. Due to the coarse-grained nature, ENMs may indicate directions of possible large scale conformational transitions of a biomolecule. In fact, often very significant overlap of the softest ENM modes with experimentally observed conformational changes in proteins has been found. This has, for example, been used in efficient flexible docking simulations of proteins [127]. The idea of coupling ENM analysis of proteins and atomistic MD simulations has been explored by Zhang et al. [233] employing a separate temperature coupling of collective ENM degrees of freedom of a molecule and temperature control of the rest of the system in a single simulation. The motion along ENM degrees of freedom is amplified by increasing the temperature álongt́he collective degrees of freedom of the molecule. The method allowed enhanced sampling of peptide and protein motion [233]. However, separate temperature coupling of different degrees of freedom corresponds to a non-physical simulation system and it is not clear if such a simulation produces conformations compatible with the desired simulation ensemble (e.g. a canonical ensemble). In addition, extended simulation runs with an increased temperature of the soft collective degrees of freedom of a system may lead to sampling of undesired conformations, e.g. unfolding of the protein (if the temperature of a collective degree of freedom is kept above the folding temperature of the protein). It is also possible to use random deformations in collective directions derived from an ENM analysis to start atomistic MD simulations and in that ways improve the conformational search properties of MD simulations. Such approach has been employed by Gur et al. [60] to study global motions in adenylate kinase. However, such approach does not allow for a canonical sampling of relevant conformational states during the atomistic simulations. In a recently developed H-REMD method different levels of a biasing potential derived from a low-resolution ENM description were used to improve the sampling of large scale motions in proteins [229] in implicit solvent. The biasing potentials act on distances between C_α backbone atoms and are

compatible with soft collective degrees of freedom of the protein derived from the ENM analysis. Similar to a flooding potential [59] the biasing potentials promote transitions along the globally soft directions in the protein system. Embedding these potentials in a replica exchange scheme can also improve the sampling in a reference replica under the control of the original force field. The method proved successful to sample global domain motions motions of the T4 lysozyme as well as to speed up folding of a peptide compared to cMD simulations [229]. However, for an efficient sampling the approach required the frequent recalculation of the ENM of the protein under study in order to adapt the biasing potential to the changes in the protein structure. As a consequence the biasing potentials are not constant over time which may result in a non-canonical sampling not only in the replicas but potentially also in the reference replica simulation. Aim of the current study is to extend the method by reducing the number of biasing potentials acting not on distances between C_α atoms but between a modest number of centroids reflecting the local atom density or domain structure of the protein. The width and strength of biasing potentials between the centroids is still determined by the ENM fluctuations and integrated into all atom H-REMD simulations of the protein in explicit solvent. In addition, dynamical adjustment of the biasing potentials during an equilibration phase can also speed up conformational sampling [153]. The performance of the method was tested on two biologically important proteins. One example was a substructure of glycoprotein 130 (gp130) [18]]consisting of the two subdomains D2 and D3. The activation of the gp130 signal transduction protein is coupled to domain-domain motions [183]. The second example was the cyanovirin-N dimer, a potent virocidal protein [13, 228, 9, 15]. Nuclear magnetic resonance [13] and crystallographic studies [228] indicate different domain dimer arrangements compatible with global domain flexibility of the system [9, 15]. The application of the ENM coupled REMD approach resulted in both cases in increased domain-domain motion compared to standard MD simulations including large scale conformational transitions compatible with available experimental data. The approach could be a valuable tool to systematically study large scale motions in biomolecules under realistic simulation conditions including explicit solvent and ions.

4.2 ENM coupled replica exchange simulations

4.2.1 MD simulations

All simulations were performed by means of the pmemd module of the Amber12 package [26]. The crystal structures pdb 1BQU served as start structures for molecular dynamics (MD) simulations of the domain 2 (D2) and domain 3 (D3) substructure of glycoprotein 130 (gp130). In case the Cyanovirin-N domain dimer, two different experimental start structures, pdb1L5B and pdb1L5E, that differed in the global domain geometry were used. The structures were solvated in truncated octahedral boxes with explicit TIP3P water molecules and neutralized with sodium and chloride ions [84] by means of the leap module and using the parm10 force field [43]. During 100 ps equilibration the systems were heated up to 298 K while protein heavy atoms were harmonically restrained (25 kcal/mol $\AA^{-2}$) to positions in the starting structure. After removing positional restraints gradually during another 100 ps the structures were

equilibrated for 10 ns (temperature: 298 K, pressure: 1 bar) and an unrestrained simulation (1 ns) served to generate 12 starting snapshots (equi-distant along the 1 ns) for the ENM coupled REMD and independent cMD simulations. Long range electrostatic interactions were calculated with the Particle Mesh Ewald (PME) method [35] and using a real space cutoff radius of 9$\AA$. Bonds involving hydrogen atoms were constrained by the Settle algorithm [133], which allowed for a time step of 2 fs. Simulation results were analyzed by means of the cpptraj module of Amber12.

4.2.2 Selection of domain centroids and calculation of fluctuations using an elastic network model

Domain centroids were selected based on the density of atoms in different protein regions. This was achieved by defining a number of centroids and minimizing the sum of distances between position of centroids and nearby protein atoms (weighted by a Gaussian function) and maximizing the distances [65]. The calculation of eigenvectors (normal modes) and eigenvalues provided orthogonal directions of motion and corresponding force constants that are inversely proportional to the fluctuation in each normal mode direction. The excitation of the 50 softest normal modes (by a thermal energy of RT, with R: gas constant and temperature T,=300 K) was used to calculate the expected distance fluctuations between the predetermined domain centroids based on the ENM analysis. The 50 softest modes are responsible for >90% of the

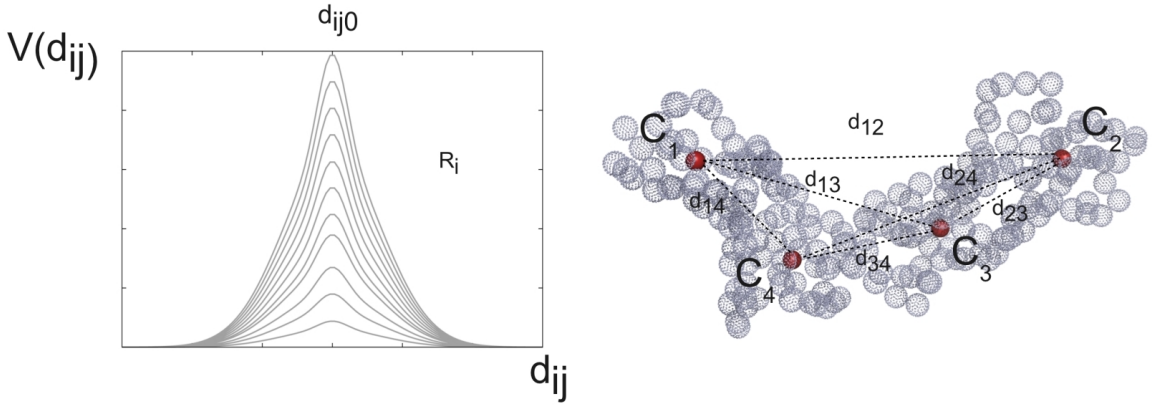


Figure 4.1: Form of the biasing potential (left panel) acting between distances (d_{ij}) of protein domain centroids (indicated as C1-C4 in the right panel). In the right panel the structure of the gpw130-D2-D3 protein is indicated by the $C_\alpha$backbone atoms (grey spheres). The biasing potentials act with increasing amplitude (indicated in the left panel) in the replica runs. One reference replica runs only under the control of the original force field without any biasing potential. The width of each biasing potential is based on the distance fluctuation between corresponding centroids calculated from an elastic network model (ENM) of the protein (see Methods for details). The domain centroids represent rigid (dense) parts of the protein (with little internal fluctuation according to the ENM).

gp130-D2-D3 ENM-REMD simulations	$< d_{120} >$	$< d_{130} >$	$< d_{140} >$	$< d_{230} >$	$< d_{240} >$	$< d_{340} >$
mean centroid-centroid-distance (Å)	20.6±0.6	38.9±0.7	23.1±0.8	55.6±0.7	43.3±0.6	20.7±0.33
mean width of 1st biasing potential (Å)	1.75±0.4	3.4±0.5	2.25±0.2	6.6±0.6	3.0±0.1	2.2±0.3
mean width of 2nd biasing potential (Å)	5.3±0.7	10.3±0.9	6.8±0.3	20.0±1.2	8.9±0.4	6.5±0.6
Cyanovirin-A ENM-REMD simulations						
mean centroid-centroid-distance (Å)	31.8±0.7	29.4±0.5	23.6±0.4	23.6±0.5	25.8±1.0	31.8±0.6
mean width of 1st biasing potential (Å)	5.75±0.45	3.6±0.5	2.3±0.3	2.2±0.3	10.1±1.3	
mean width of 2nd biasing potential (Å)	17.1±0.8	10.7±0.9	6.8±0.7	6.7±0.6	30.4±2.4	17.2±0.9
Cyanovirin-B ENM-REMD simulations						
mean centroid-centroid-distance (Å)	35.7±0.6	27.4±0.5	23.1±0.4	23.2±0.5	41.2±0.9	36.2±0.4
mean width of 1st biasing potential (Å)	6.3±0.4	7.1±0.4	2.0±0.2	2.0±0.3	14.2±0.7	6.7±0.4
mean width of 2nd biasing potential (Å)	18.8±0.7	21.4±0.7	6.1±0.5	6.1±0.6	42.6±1.4	20.1±0.6

Table 4.1: The centroid-centroid distances (d_{ij0}) correspond to the reference distances in the biasing potentials in each replica and the mean width determine the width of the biasing potentials (see equation in Methods section and Fig. 4.1). The mean increase in amplitude of the biasing potential along the replica ladder was 0.082±0.02 kcal/mol (gp130-D2-D3 system), 0.18±0.045 kcal/mol (cyanovirin-A), and 0.18±0.04 kcal/mol (cyanovirin-B)

conformational fluctuations of the protein systems (not shown). A distance-dependent biasing potential was designed in order to specifically promote structural changes in the atomistic simulations compatible with the distance fluctuation estimates from the ENM and implemented in the pmemd module to destabilize the protein structure along the centroid distances d_{ij} (i,j are centroid labels).

$$V(d_{ij}) = k([d_{ij} - d_{ij0}]^2 - \Delta d_{ij}^2)^2, if |d_{ij} - d_{ij0}| \leq \Delta d_{ij} V(d_{ij}) = 0, otherwise \qquad (4.1)$$

The biasing potential is characterized by a positive (repulsive) amplitude and a characteristic width similar to a Gaussian potential. For each centroid pair it was centered to a reference centroid distance (d_{ij0}) and using a width proportional to the precalculated distance fluctuations from the ENM analysis. The effect of such a potential in the atomistic replicas runs is an increased motion along the centroid distance directions compatible with the ENM calculations. For each centroid-centroid distance two such potentials were used one with a narrow width and one with broader range (see Fig. 4.1, Table 4.1).

4.2.3 Replica exchange scheme with adaptive biasing levels

The biasing potential based on the ENM calculations of the protein molecules was integrated into H-REMD simulations to improve conformational sampling at the level of atomistic explicit solvent simulations (illustrated in Fig. 4.2). During the H-REMD simulations one reference replica was propagated under the control of the original parm10 force field whereas the centroid-centroid distance dependent biasing potentials were added with increasing amplitudes in each of the 11 replicas (total number of replicas was in each case 12 including the reference simulation). A replica exchange was attempted every 2 ps and accepted or rejected according to the standard Metropolis criterion in H-REMD simulations[152]. The magnitudes and the width of the biasing

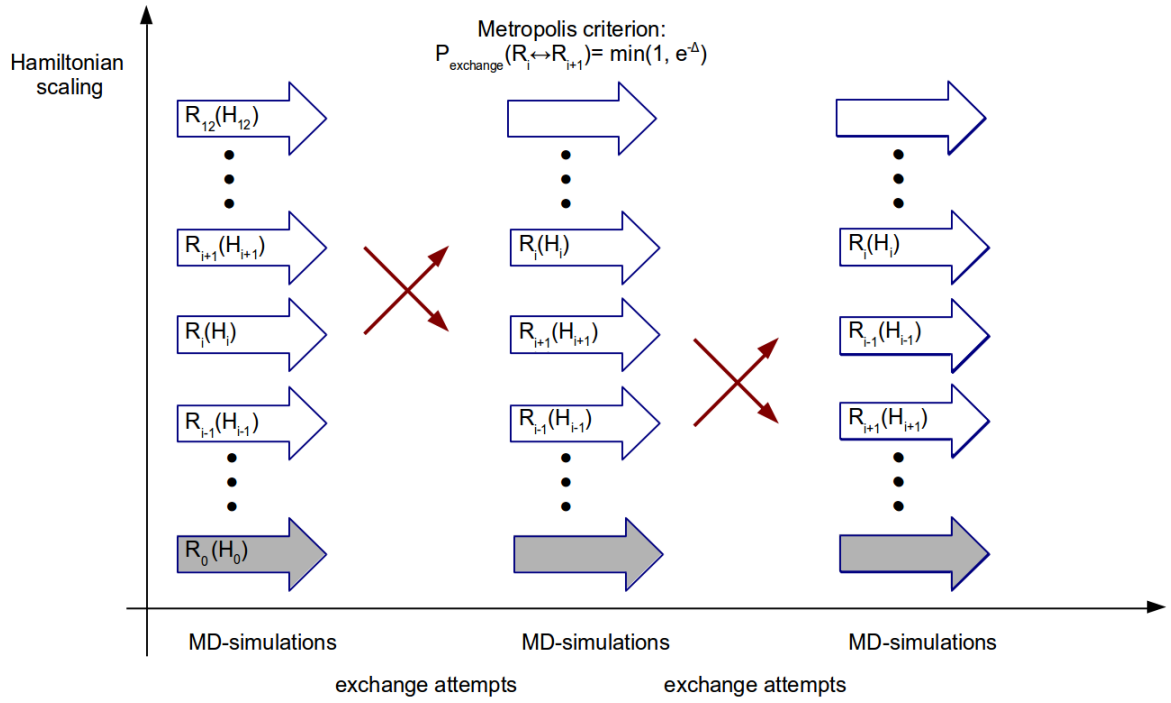


Figure 4.2: Illustration of the ENM-coupled replica exchange simulations. Exchanges between replicas (left panel) were attempted every 2 ps. Every 100 exchange attempts (200ps), the centroid-centroid distances and ENM-derived fluctuations to determine the width of the biasing potential (illustrated in the right panel) were recalculated from the running average over the past 200 exchange attempts. In addition, the amplitude of the biasing potential used in the replicas is adjusted to keep a high exchange acceptance rate based on the exchange acceptance rate of the past 500 exchange attempts. The shape of the biasing potential and the distances between domain centroids (for an arbitrary centroid arrangement) are indicated in the left panel.

potentials in the replicas were adjusted during the simulations (Fig. 4.2)). The reference centroid-centroid distances were recalculated every 200 ps from the running average of the last 400 ps. Similarly, the magnitude of the biasing potentials was also adjusted every 200 ps to optimize the acceptance rate of replica exchanges. Test calculations indicated an average over the last 500 exchanges to provide for a reasonable estimation of the acceptance rates. Based on a previous adaptive H-REMD approach [153] the amplitude of the biasing potentials was lowered (increased) by 10 % if the acceptance probability for exchanges between neighbors fell below 20% (surpasses 60%) in any of the replicas. The biasing levels were also lowered by 10% if a configuration visits each replica less than once within the previous 500 exchanges. The last criterion promotes that the replica conformations do not only exchange with directly adjacent neighboring replica conditions but reach all available biasing levels from time to time. The choices of adjustment resulted in smooth adaptation of the biasing levels in each replica to near constant levels at the final stage of the simulations (table 4.1).

4.3 Conformational flexibility of the D2-D3 interdomain connectivity of glycoprotein 130

The glycoprotein 130 (gp130) is a membrane-bound cytokine receptor involved in signal transduction events during inflammation processes. It consists of six domains (D1-D6) and forms complexes with interleukin 6 (Il-6) [63, 126]. Il-6 binds to the D2-D3 domain hinge region which mediates subsequent activation of inflammatory processes [63, 126]. Experimental crystal structures of several gp130 sub-domains in the absence and presence of bound Il-6 are available. The D2-D3 segment forms an independent stable unit that was crystallized in the absence and presence of Il-6 [18]. Recent electron microscopy studies on a closely related system [125] and mutagenesis of the D2-D3 hinge region in combination with molecular dynamics (MD) studies [183] indicated that changes in the D2-D3 inter domain angle are involved in the gp130 activation process. Inactive in its kinked crystal-like form, it is found to be active in a more extended form [183]. Usually induced by Il-6 binding, this stretching process of gp130 can be alternatively provoked also by mutations at the D2-D3 domain interface [183]. In order to test the ability of the present ENM-coupled H-REMD (ENM-REMD) approach to enhance global domain motion and to explore the domain mobility on the gp130-D2-D3 system we performed both classical continuous (c)MD and ENM-REMD simulations. Conformations changed and fluctuated only modestly in 12 independent 20 ns cMD simulations of the gp130-D2-D3 segment each starting from different initial conditions. This is reflected in a root mean square deviation of the protein backbone (RMSD) below 3 $\AA$ during all 12 simulations (Fig. 4.3A)). In contrast, during ENM-REMD simulations much larger conformational changes and fluctuations were found in the reference replica controlled by the original force field. For a small fraction of sampled states in the reference replica the RMSD reached up to ca. 11-12 $\AA$ from the starting structure (Fig. 4.3A)). The sampled domain arrangements can be characterized by an angle ϕ and a dihedral angleθ defined for the protein molecule (Fig. 4.3)D).A projection of the sampled conformations during all combined cMD simulations (240 ns) indicates only a modest sampling of D2-D3 interdomain angles (between 125-140^o) similar to previous cMD simulations on the same system [183] and a narrow sampled regime of the pseudo dihedral angle θ (Fig. 4.3C). For the ENM-REMD simulations (only considering the 16 ns of the reference replica) a larger range of domain-domain conformations especially with larger interdomain angle and a much broader range of pseudo dihedral angles were observed (Fig. 4.3B,C). This indicates that the gp130-D2-D3 segment can adopt a greater variety of domain arrangements then just the crystal structure conformation. However, the percentage of these alternative arrangements (ca. 5-10%) is lower than the arrangement observed in the crystal structure. There is experimental evidence from cryoEM studies on the full gp130 structure that indeed the D2-D3 domain segment can undergo significant conformational changes with different interdomain angles [125]. It is interesting to compare the results of the coupled ENM-REMD simulations with an ENM analysis of the gp130-D2-D3 global mobility. The deformations in the first and second softest deformation modes (Fig. 4.4) clearly indicate significant possible mobility in the domain angle coordinate (first mode) and

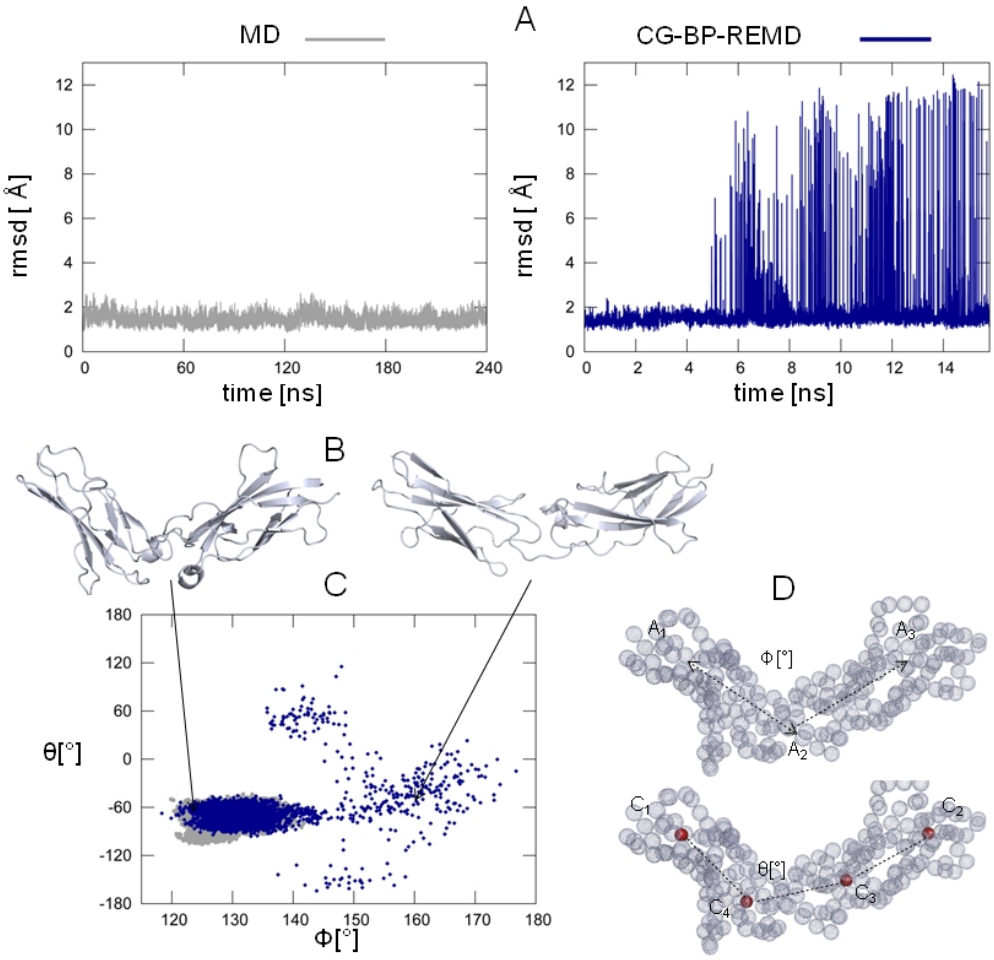


Figure 4.3: (A, left panel) Root mean square deviation (RMSD) of the gp130-D2-D3 backbone from the crystal structure during 12x20 ns cMD simulations. Note, that the recorded RMSD of 12 independent simulations (each 20 ns) were appended and shown in one plot. (A, right panel) RMSD vs. simulation time observed in the reference replica (without biasing potential) of the ENM-REMD simulation starting from the same gp130-D2-D3 start structures as the independent cMD simulations. (B) Snap shots (cartoon representation of a slightly more kinked gpw130-D2-D3 domain arrangement (left) and a more straight gpw130-D2-D3 domain arrangement compared to the crystal structure observed during the ENM-REMD simulations. The positions of these structures among sampled states projected onto a domain angleΦ vs. domain dihedral angle Θ plot (Θ -Φ plot) is indicated as arrows in (C). Each dot in (C) corresponds to a sampled domain arrangement in the ENM-REMD (blue) or cMD simulations (grey dots). The definition of the angular and dihedral angle variables to approximately describe the domain arrangement is indicated in (D). While angleΦ is spanned by centers of masses of domain 2(A1 res 1-93), the hinge (A2 residue 90-99) and domain 3 (A3 residue 97-201), dihedral angle Θ is defined by the 4 centroids of the starting structure (C1-C4)

dihedral angle coordinate (second mode). Indeed, as discussed above, the ENM coupled REMD also show significant mobility in these variables. This indicates that the

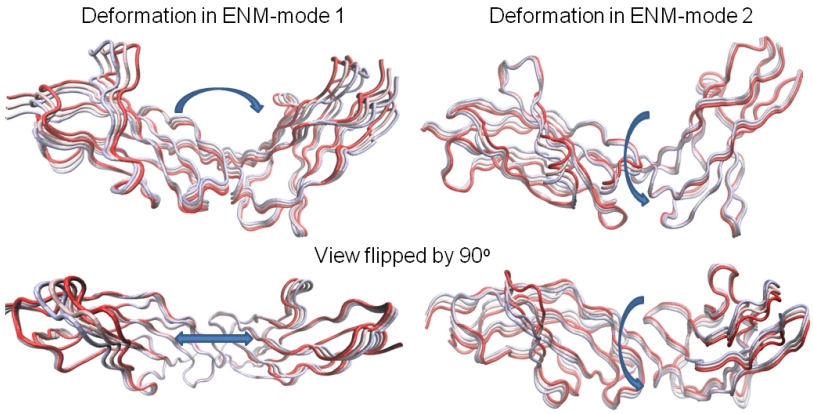


Figure 4.4: Deformation of gpw130-D2-D3 in the two softest normal modes calculated by an elastic network model. Structural deformations (red and grey tubes) were generated by displacing the crystal structure of the gpw130-D2-D3 protein (pink tube) in negative and positive direction along the softest (left panels) and second softest ENM-mode (right panel). The softest modes are characterized by the smallest associated eigenvalue or force constant for deformation and are predicted to contribute most to the fluctuation of the protein structure.

ENM calculation provides collective degrees of freedom that correspond to directions of global motions also seen in the atomistic simulations (in the reference replica that runs without any biasing). However, the atomistic REMD simulations do not result in a simple Gaussian distribution of sampled states in these global variables (as an ENM model would suggest) but show a much more complex distribution. For example, domain angular variable distributes very asymmetrically including mainly a set of more open extended states. The resulting non-uniform distribution of states in the plain spanned by the angular and dihedral variables (Fig. 4.3C) indicates free energy barriers between different globally distinct arrangements. These barriers prevent a sampling of alternative global states during the 240 ns cMD simulations but in the ENM-REMD the biasing potentials in each replica help the system to overcome such energy barriers.

4.4 Domain mobility of cyanovirin -N

The cyanovirin-N protein from cyanobacterium Nostoc ellipsosporum is a highly potent virucidal agent that has found interest as a potent HIV (human-immuno-deficiency virus) inactivating protein [15]. It was found to form a metastable (domain-swapped) dimer in solution and in crystal structures [9, 15] and can form different inter-domain arrangements depending on the experimental conditions. Two crystal structures pdb1l5e and pdb1l5b obtained under different crystallization conditions, termed cyanovirin-A (pdb1l5b) and cyanovirin-B (pdb1l5e), served as starting structures for comparative cMD and ENM-REMD simulations. The two crystal structures differ in the interdomain geometry (Fig. 4.5) giving rise to an overall backbone RMSD of 7 $\AA$. In these proteins residues 1-50 and 151-202 as well as 51-151 form stable rigid domains. Each of the stable domains can be superimposed with an RMSD of < 1 $\AA$. After superposition

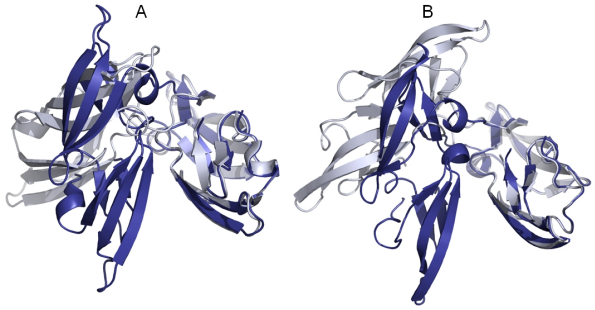


Figure 4.5: Comparison of cyanovirin conformation A (blue) and conformation B (grey) before (A) and after (B) equilibration. The structures were superimposed with respect to the first domain dimer (right side in each panel, corresponding to residues 1-51 and 151-202) and only the backbone is illustrated (cartoon representation). During equilibration a slight shift of the domain arrangements compared to the start geometries was observed resulting in total backbone RMSD deviations from the starting structure of 1.5 $\AA$ (cyanovirin-A) and 4.5 $\AA$ (cyanovirin-B), respectively. The structures of individual domains in each domain dimer did not change significantly (backbone RMSD with respect to start structure was ł 1 $\AA$). The total RMSD between cyanovirin-A and B conformers was similar for the starting structures (7.3 $\AA$, shown in A) and equilibrated conformers (8.0 $\AA$, shown in B).

of the two crystal structures with respect to such a domain the RMSD of the other domains reach ca. 17.5 $\AA$ indicating that the large total RMSD of 7 $\AA$ between the two crystal structures is mainly due to a domain-domain rearrangement. After an 10 ns cMD equilibration run conformations of cyanovirin-A and cyanovirin-B shifted resulting in a ca. 1.5 $\AA$ and ca. 4.5 $\AA$ RMSD from the respective crystal start structure. Despite these shifts the overall dimer domain arrangement was conserved (Fig. 4.5) and the total RMSD between the equilibrated structures cyanovirin-A/B was ca. 8 $\AA$. The result indicates a significant flexibility of the structures in the proximity of the corresponding starting geometry. Starting from the equilibrated cyanovirin-A structure during 12 independent cMD simulations conformational deviations from the equilibrated starting structure of up to 4 $\AA$ were observed (similar to the large fluctuations observed in the equilibration phase). However, no transitions to conformations close to the alternative domain arrangement (cyanovirin-B) were found (Fig. 4.6). In contrast, several conformations switched from the starting state to the alternative domain arrangement in the ENM-REMD reference replica (corresponding to grey points in the upper panels of Fig. 4.6 with a small RMSD with respect to structure of cyanovirin-B). In case of starting from the equilibrated cyanovirin-B arrangement the cMD simulations indicate a transition into the alternative cyanovirin-A in 1 out 12 independent 20 ns simulations (Fig. 4.6). This transition was found to be irreversible on the time scale of the cMD simulations. Cyanovirin-B switched towards the cyanovirin-A domain topology in the reference replica of the ENM-REMD simulations although less frequently compared to the opposite transition. For fully equilibrated simulations the distributions of the two domain arrangements in the reference replicas of both simulations (starting from different domain arrangements) are expected to be identical. Such equilibration of states is not reached within the 20 ns REMD simulation. However, a clear trend can

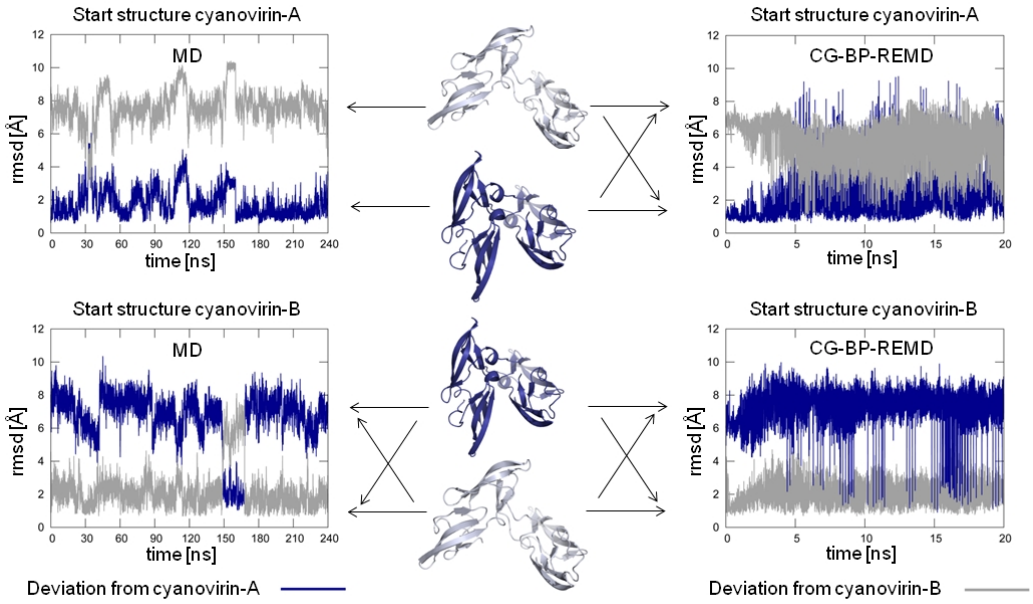


Figure 4.6: Upper left panels): Backbone RMSD from equilibrated cyanovirin-A structure during 12 cMD simulations (results of 12x20 ns concatenated in the upper left graph). Both the deviation from the equilibrated cyanovirin-A (blue line) and cyanovirin-B conformer (grey line) are plotted. (Upper right panel): Same for the reference replica of an ENM-REMD simulation. In the middle panel representative conformations for cyanovirin-A (blue cartoon) and B (grey) conformers to indicate the cyanovirin structure corresponding approximately to the RMSD level in each plot. (Lower panels): Same results starting from the equilibrated cyanovirin-B conformer (color coding is the same as in the upper panels).

be seen reflected in an increasing accumulation of transitions to the cyanovirin-B type arrangements in the simulation starting form cyanovirin-A and only little sampling of the alternative cyanovirin-A if simulations were started from cyanovirin-B (Fig. 4.6). This result indicates the cyanovirin-A conformer to be presumably more stable in the current force field than conformer cyanovirin-B. Although alternative domain arrangements were not sampled in full equilibration, frequent domain-arrangement transitions were observed in the ENM-REMD reference replica: In contrast only one transition occurred in cMD simulations of same total length. Thus, the performance of the ENM-REMD approach in terms of sampling alternative global arrangements is also much better for the second example compared to cMD simulations at the same computational demand.

4.5 Conclusions

Elastic network models often allow a rapid estimation of the collective directions for global domain motions in proteins. However, even along these soft collective directions the protein system might have to cross numerous local minima and barriers at an atomic resolution description in order to reach alternative stable arrangements during

available simulation time. There have been several attempts to combine ENM analysis with atomistic MD simulations. Among these are approaches to initially deform structures along ENM soft degrees of freedom and use the generated structures to start MD simulations at atomic resolution [60]. Alternatively, separate temperature coupling along collective degrees of freedom (with increased associated thermal energy) either in single MD simulations [233] or coupled with temperature (T-)REMD [102] have been described. These approaches may improve the conformational search properties but can strongly disturb the equilibrium of an MD-simulation and may result in a largely non-canonical sampling of available states. A major aim of advanced sampling schemes is to produce the same distribution of states in shorter time as long continuous MD simulations. This is especially important in case of global domain motions where different domain arrangements are often populated to similar degrees corresponding to small free energy differences. Even a small non-canonical bias in the sampling techniques may then strongly affect the population of alternative structures. In a previous study we used a biasing potential acting on many C_α-C_α backbone distances in the protein system compatible with the distance fluctuations derived from an ENM analysis [229]. The frequent recalculation of the ENM during the simulations resulted in biasing potentials that changed and fluctuated during the simulations. Such varying biasing potentials can also influence the sampling in the reference replica and may cause deviations from canonical sampling. In the present technique only a few biasing potentials acting on the distances between centers-of-mass of protein segments are used. The position and distances between these centers change less and much more slowly than individual atom distances during the simulations. In addition, for the recalculation of biasing potentials based on the ENM we use the average over a large simulation window instead of an instantaneous structure (as in the previous method). This procedure resulted in biasing potentials that adapt during the simulation time towards stable potentials in each replica with small residual fluctuations (see table 4.1). Hence, the application of the present REMD with an ENM derived biasing potential in the replica runs allows an enhanced domain-domain motion but at the same time a canonical sampling in the reference replica compatible with the original force field and free of any biasing. In addition, based on previous experience [153] the adjustment of the biasing potential amplitudes in each replica guarantees high acceptance rates and high conformer turn around in the space of replica runs. For the two test systems gp-130-D2-D3 and cyanovirin-N the ENM-REMD approach enhanced sampling of different domain arrangements compatible with available experimental data could be demonstrated. Although on the present time scale no converged sampling could be achieved the method showed clearly better sampling of cMD simulations of total length identical to the combined REMD replica runs. In future efforts different forms of the biasing potentials will be tested. However, even in the present form the ENM-REMD technique could be useful to study systematically domain-domain motions in proteins under realistic simulation conditions including surrounding solvent and ions.

Chapter 5

Allosteric coupling between side chain interactions and binding pocket flexibility in HLA-B*44:02 molecules

As part of the viral immune response MHC class I molecules present antigenic peptides to cytotoxic T-cells at the cell surface. This chapter centers around two human class I allotypes HLA-B*44:05 and HLA-B*B*44:02, whose sequence is identical except for a single amino acid at position 116 (i.e tyrosine 116 B*44:05 aspartic acid B*44:02) resulting in very similar structures in complex with the same peptide. In contrast to B*44:05 peptides bind efficiently to B*44:02 in the presence of the chaperone protein tapasin. Binding site and role of tapasin during the loading process still are elusive. However, the conformational variety of B*4402 in its peptide free form gives insight into the most flexible parts of proteins, that may be stabilized by tapasin. The biasing potential replica exchange method described in chapter 3 promotes side chain flips of selected amino acids in MHC class I molecules. While the rotameric state of the basic Arg 97 turns out to be coupled to the global shape of peptide free B*44:02 it does not affect the form of peptide free B*44:05. Conformational flexibility becomes manifest in the the F pocket that is much more mobile in the absence of a peptide in case of B*44:02 compared to B*44:05. Free energy simulations to open the F pocket indicate a molecular side chain switch mechanism that underlies the global opening motion. This side chain switch involves the rearrangement of salt bridges and hydrogen bonding of the basic Arg 97 with three acidic aspartate residues 114, 116 and 156 near the F pocket. Rotations of the side chains of these residues are accelerated with the biasing replica exchange method: Same side chain rearrangements prove to induce global opening motions of the F pocket for the B*44:02 in contrast to B*44:05 allotype.

5.1 Motivation and outline

MHC class I molecules play a key role in the antiviral human immune response. In the endoplasmic reticulum (ER), class I molecules form complexes with short antigenic peptides, assisted by chaperone proteins [151]. After successful peptide loading, class I molecules are transported to the cell surface to be presented to the immune system which can lead to destruction of infected cells by cytotoxic T cells [78, 139]. Antigenic peptides are bound to MHC class I molecules in a narrow cleft on top of an extended beta sheet flanked by two alpha helices [2] Fig. 5.1A). Among the products of the three class I gene loci in humans (HLA-A, B, C), HLA-A and HLA-B play the most prominent role in presenting antigenic peptides at the cell surface and show the greatest polymorphic variation. The latter enables HLA-A and HLA-B molecules to bind to a large number of different peptides (every allotype binds to a specific set; [232]). Despite the large polymorphic variance among HLA-A/B molecules, the basic concept of peptide binding is well conserved [54, 124, 16, 2]. By forming transient complexes with class I molecules in the ER, the chaperone protein tapasin assists peptide loading and promotes preferential binding of high-affinity peptides in the editing process [179, 224, 182, 77]. Mutagenesis experiments on class I molecules and tapasin indicate transient binding of tapasin to a helical segment (α_{2-1} helix, Fig. 5.1B) and to residues 221 and 221 in the α_3 domain of the class I molecules [73, 111, 148, 202, 212, 219, 231]. Although most class I molecules benefit from the presence of tapasin in the ER, there are some

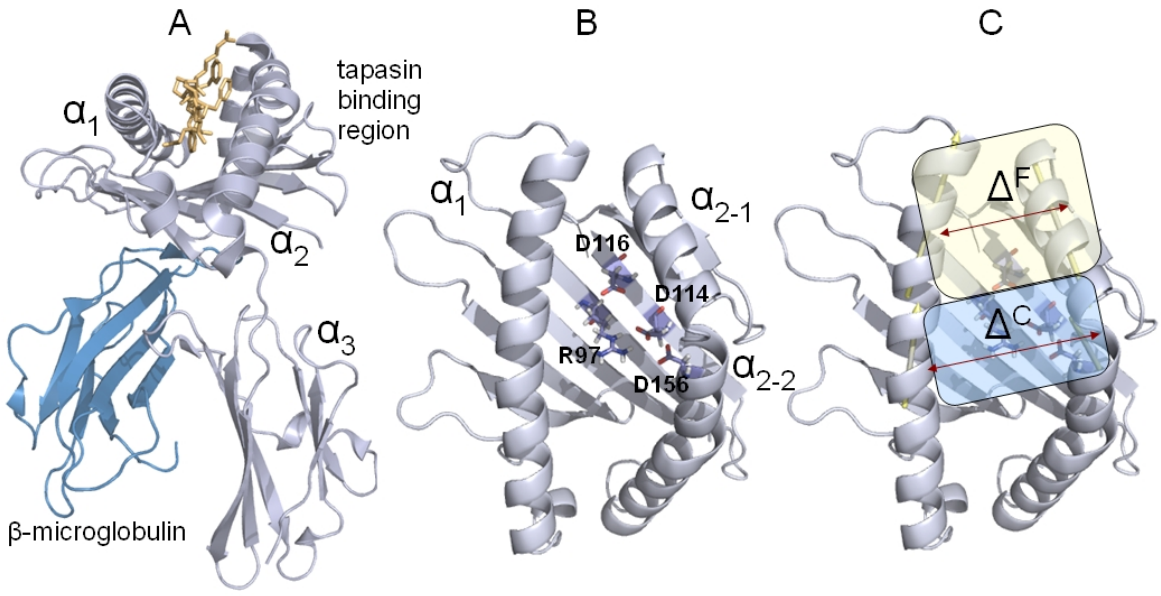


Figure 5.1: (A) Peptide (yellow sticks) loaded MHC class I molecules with $\alpha_{1,2,3}$ domains and non-covalently attached $\beta_2 m$ (cyan) in cartoon representation. (B) Enlargement of the B* 44:02 binding pocket (rotated by 90° with respect to view in A) with the R97 side chain at the floor of the peptide binding cleft located close to three acidic aspartic acids, D114, D116, and D156 (labeled residues). (C) Location of C and F pockets and definition of distances (red double arrows) for measuring pocket width (centers of mass of the α_{2-1} helix and α_1 helix segment in case of the F pocket).

allotypes that show reduced dependence or even complete tapasin independence for cell surface expression [221, 154, 231]. Interestingly, single amino acid differences between two allotypes can have a major impact on the tapasin dependence of class I allotypes [221]. A particularly interesting example is the allotype HLA-B*44:05 that was shown to form stable complexes with antigenic peptides and exhibits cell surface presentation in cells that lack tapasin [221, 231]. In contrast, under the same conditions, cells with the closely related allotype HLA-B*44:02 showed strongly reduced presentation of peptides at the cell surface (in the absence of tapasin). Both allotypes differ only by a single residue at position 116 (a tyrosine in case of HLA-B*44:05 and an aspartate in case of HLA-B*44:02, respectively) at the bottom of the F pocket region that binds the peptide C-terminus (Fig. 5.1B) and are not located close to the proposed tapasin binding region of class I. Crystal structures of both allotypes in complex with the same high-affinity peptide do not show any difference that explains differences in tapasin dependence [55, 122, 230, 231]. Thus, residue 116 may have an effect on the structure and dynamics of peptide free class I molecules, which influences the interaction and recognition by tapasin. Buried at the interface to the peptide, residue 116 is also inaccessible to tapasin in both cases. So far, experimental studies based on NMR spectroscopy indicate that MHC class I molecules are at least partially folded in the absence of antigenic peptides [17, 105]. In a previous comparative molecular dynamics (MD) simulation study, we found especially the C-terminal part of the α_2 helix (termed α_{2-1} helix, residues: 135-147) showing enhanced flexibility and conformational shifts in the absence of a bound peptide in case of the B*44:02 allotype [189]. Indeed, this helical segment proved also to be most mobile in a comparative analysis of several crystal structures of class I molecules [111, 47]. Strikingly, the mobile α_{2-1} helix corresponds to the proposed tapasin binding region described above. In contrast to B*44:02, the same α_{2-1} helix segment of peptide-free B*44:05 fluctuates closely around the peptide-loaded X-ray structure (receptive state) in MD simulations [189]. This suggests that tapasin may stabilize empty allotypes (with a mobile α_{2-1} helix) that subsequently adopt a receptive conformation [182, 55]. Interestingly, simulations predicted a D116H mutation of B*44:02 to reduce the α_{2-1} helix mobility of peptide-free molecules to a similar level as observed for B*44:05. Hence, this mutation should result in a tapasin-independent allele, which was indeed confirmed by experiments [55]. While these studies indicate a coupling between the mobility of the F pocket (in the absence of a peptide), in particular of the $\alpha_2 - 1$ helix and the tapasin dependence of peptide loading, the molecular origin of the enhanced F pocket mobility is not yet clear. It is mechanistically interesting how one single residue difference in the F pocket floor (D116 in B*44:02 vs. Y116 in B*44:05) can affect the global dynamics of an entire binding region. In the present study, we have investigated the coupling of local side chain motion in the binding cleft and global mobility of the α_{2-1} helix segment of the B*44:02 allele by means of umbrella sampling (US) and biasing potential replica exchange (BP-REMD) simulations. We find a side chain molecular switch mechanism of the R97 residue: Located at the floor of the binding cleft, the R97 side chain can interact with different acidic residues in the F pocket (including residue D116). Simulations indicate a causal relation between different interaction networks of the R97 side chain and global shifts of the α_{2-1} helix segment. Induced global changes of the α_{2-1} helix segment resulted in a switch of the R97 side chain interactions (US simulations). On the other hand, induced switching between alternative R97 side chain interactions

led to a global shift of the α_{2-1} helix segment (BP-REMD-simulations). Although the mechanism is specific for the B*44:02 allotype similar types of correlated side chain rearrangements may cause global changes in other allotypes and may correspond to a general mechanism valid in other protein molecules as well.

5.2 Molecular dynamics simulations and biasing potential replica exchange MD simulations

The crystal structures of human class I B*44:02 allotypes ([122]) protein data bank entry: pdb1M6O and B*44:05 ([231] protein data bank entry: pdb1SYV) in complex with the same peptide EEGRAFSF served as start structures for all Molecular Dynamics (MD) simulations. To model an initial structure of the empty B*44:02 and B*44:05 allotypes, respectively, the peptides were removed from the crystal structures (Fig. 5.1A,B). Centered in a truncated octahedral box solvated with TIP3P waters [84]. each protein was neutralized with Na+ and Cl− ions up to a final concentration of ca. 0.15 M. All Simulations were performed with the Amber12 package [26] by means of the parm03 force field [43]. Long range electrostatic interactions were included by means of the particle mesh Ewald option [35]. After energy minimization of each system, the protein and solvent were heated in 100 K steps (each for 0.05 ns), including positional restraints on all heavy atoms (50 kcal/(mol$\mathring{A}^2$). At 300K, restraints were removed gradually during additional 0.1 ns simulation time. Opening of the binding pocket of peptide free B*44:02 was enforced by umbrella sampling simulations: A harmonic potential with a restraining force constant of 2.0 kcal/(mol$\mathring{A}^2$) was imposed on the heavy atoms of two alpha helices that surround the F pocket floor (see Fig. 5.1C, residues 74-85 and 138-149). Applying these distance restraints in a range from 10 to 26 $\mathring{A}$ in 0.5 $\mathring{A}$ steps, each umbrella window was simulated for 9 ns. Including data of last 4 ns of these runs, the resulting distance histograms were transformed into a free energy profile by means of the weighted histogram analysis method (WHAM [104]; implemented by Grossfield, http://dasher.wustl.edu/alan/). A biasing potential Hamiltonian replica exchange (BP-REMD) method proved successful to disturb selected key side chains in the F pocket locally. The simulations followed the protocol described in [153] without biasing the backbone conformation. Side chain flips of the amino acid R97 and adjacent residues D156 and N77 were promoted by a penalty potential, whose three plateaus with a width of 40° are centered around stable χ_1 angle regimes at -60° 60° and 180°. BP-REMD simulations consisted of a set of 5 parallel replica MD simulations, that differed by the levels of χ_1 biasing potentials: while the reference replica was under the control of the original force field without biasing potential, plateau height increased linearly for subsequent replicas. It is adjusted on the fly such that replicas travel efficiently through potential conditions and level out at 0.7 kcal/mol per side chain and replica and 0.5 kcal/mol in case of 44:05, respectively.

5.3 Global F pocket opening of HLA B*44:02 induces local R97 side chain motions

Peptide binding contributes critically to the stability of MHC class I molecules, such that MHC class I molecules are conformationally less ordered and more flexible in the empty state (without bound peptide) compared to the peptide bound state [17, 105]. Indeed, recent molecular dynamics (MD) simulation studies indicated enhanced conformational mobility especially of the α_{2-1} helix segment that flanks the F pocket in the tapasin-dependent B*44:02 allotype in the absence of a bound peptide [189]. This resulted in partial opening of the F pocket. In contrast, for the tapasin independent B*44:05 allotype which differs from B*44:02 in one residue (D116Y), only a modest increase of conformational flexibility in the absence of bound peptide compared to the bound state was observed. Hence, binding of a peptide to B*44:02 requires a substantial change in the MHC class I backbone structure (and flexibility) and presumably also of side chains near and at the F pocket. Analyzing these interactions and driving forces provides the basis for a detailed understanding of the peptide loading mechanism and the role of the tapasin chaperone protein during this process. MD simulations are especially well suited to study the global F pocket opening motion observed for B*44:02 as well as possible associated side chain motions. A convenient approach to systematically investigate the global opening of the F pocket is the application of umbrella sampling (US) simulations. By applying a penalty potential to modify the distance between centers of mass of the α_{2-1} helix segment (residues 138-149, indicated in Fig. 5.1) and a part of the α_1 helix (residues 74-85) it is possible to systematically increase the F pocket size during a series of simulations. This method was also used by Sieker et al., ([189], see Fig. 5.1C and Methods section for details). The simulations also allowed the calculation of a free energy profile for opening of the F pocket along the F pocket distance (Fig. 5.2) and gives insight into possible barriers or stable substates along the opening pathway. Similar to a previous free energy profile obtained in simulations on just the α_1/α_2 peptide binding superdomain of the B*44:02 allotype (without the α_3 domain and the $\beta_2 m$ light chain, [189]), only a small free energy difference between a near-native arrangement (F pocket distance of ∼12-13 $\AA$, termed receptive state) and a more open F pocket width (larger F pocket distance of ∼16-20 $\AA$, termed open state) were found. In contrast to the simulation on the full B*44:02 molecule, the free energy profile in case of the isolated α_1/α_2 peptide-binding

allotype	α_1-α_2-superdomain only*	entire class I structure
HLA B*4402	-2.0 kcal/mol	0.5 kcal/mol
HLA B*4405	2.1 kcal/mol	4.5 kcal/mol

Table 5.1: Free energy of F pocket opening. States with a center-of mass distance of 12-14 $\AA$ between residues 74-85 of the α_1 helix and residues 138-149 of the α_2 helix were considered as receptive or receptive conformation of the F pocket and distances between 14-22 $\AA$ as open F pocket states. Free energy differences correspond to the differences of the averages taken over the corresponding distance regimes. Free energy differences between receptive/receptive and open form are taken from Sieker et al. [189]

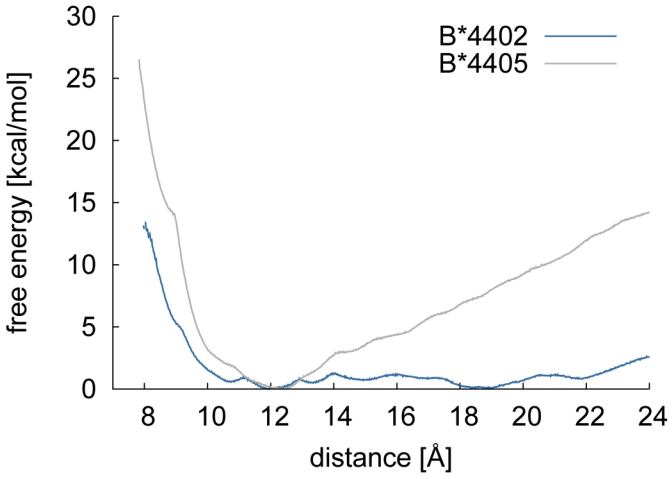


Figure 5.2: Calculated free energy change vs. F pocket width for B*44:02 and B*44:05. The free energies were extracted from umbrella sampling simulations of induced F pocket opening along the distance between $\alpha-$helices flanking the F pocket (i.e. spanned by residues 74-85 and 138-149, respectively, see Methods section for details).

superdomain predicted a lower free energy of the F pocket in open form compared to a near-native bound arrangement (table 5.1). The present umbrella sampling simulations on the full B*44:02 allotype structure resulted in a slightly lower free energy of the near-native bound arrangement of ca. 1 kcal/mol, compared to a more open arrangement (Fig. 5.2). Taken together, the results suggest that both the presence of β_2m and the α_3 domain stabilize the α_1/α_2 peptide binding domain in the peptide-receptive conformation. This result agrees with the experimental observation that indeed β_2m and peptide binding are coupled and a loss of β_2m typically destabilizes peptide binding, possibly by affecting the receptive form of the α_1/α_2 binding superdomain [210, 48]. Even though the free energy of the open form is slightly higher than in the near native bound structure, the small difference still indicates a significant population of this state in the absence of a bound peptide. For the B*4405 allotype, a significant increase of the free energy upon opening of the F pocket was observed amounting to ca. 4.5 kcal/mol. This is in qualitative agreement with previous simulations on just the α_1/α_2 peptide binding domain that showed a free energy difference between open and receptive (near-bound) F pocket distance of ca. 2.0 kcal/mol. Consistent with the observation for B*44:02, it predicts that the presence of the α_3-domain and β_2m stabilizes the α_1/α_2 peptide binding domain in a receptive conformation and thereby enhances peptide binding capacity. The large free energy barrier observed for B*44:05 suggests that open F pocket states are barely populated in peptide-free B*44:05. Interestingly, the difference between the free energy of F pocket opening in the absence vs. presence of α_3-domain and β_2m is similar for both allotypes. This suggests that in both allotypes the contribution of the α_3-domain and β_2m to the stabilization of a receptive conformation is similar. In the US simulations of the B*44:02, we observed characteristic rearrangements of the side chain structure in the F pocket region. In particular, this involved the side chain of R97 and its interaction partners. In the case of B*44:02, the basic R97 can principally interact via hydrogen-bond stabilized

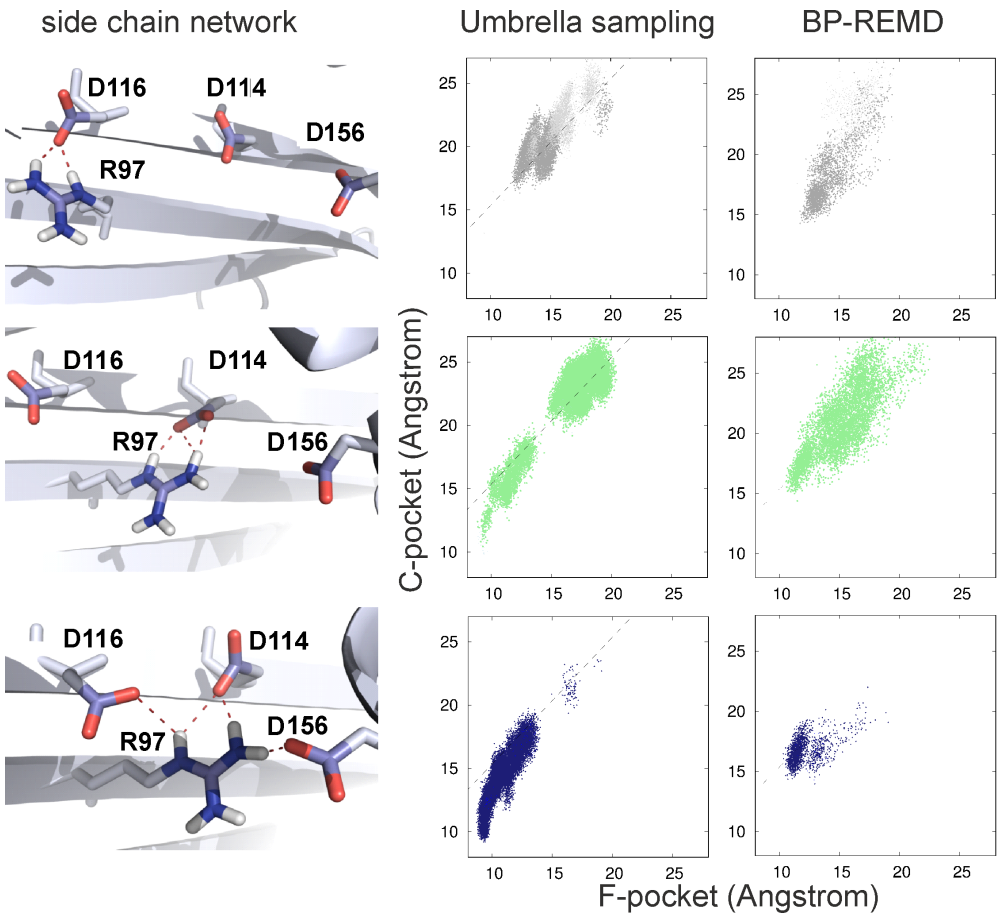


Figure 5.3: F pocket shape and characteristic side chain arrangements during induced opening and BP-REMD simulations of B*44:02. Induced opening was achieved either by umbrella sampling but also observed in BP-REMD simulations, where transition of local side chain conformations (e.g. R97) was accelerated with a biasing potential in replica runs (see Methods). The first column shows characteristic structural snapshots of open (upper panel), intermediate (middle panel), and receptive (lower panel) arrangements formed by R97 with D114, D116 and/or D156. The second and third columns indicate the corresponding distributions of the side chain states at various F pocket and C pocket distances sampled during umbrella sampling or BP-REMD simulations. The color code in columns 2 and 3 follows table 5.2 indicating sampling of receptive side chain states mainly at small C and F pocket distances (lower panels).Intermediate side chain states were found for both small and large pocket sizes whereas the open side chain arrangement was sampled only in case of large pocket width (upper panels).

salt bridges with the three nearby acidic residues D114, D116, or D156 (illustrated in Fig. 5.1 and 5.3). In the peptide-receptive conformation with a F pocket width of 12-13 $\AA$ the first are characterized by the following binding pattern: R97 is hydrogen bonded to D156 and occasionally interacting with D114 and D116. In the following, this arrangement is termed the receptive switch (see Fig. 5.3 bottom of column I and table 5.2). The criterion for a contact is fairly loose, only requiring a minimal distance (2.5 $\AA$) between any of the 5 hydrogens of the R97 guanidinium groups and

state	H-bonding R97:D156	H-bonding R97:D114	H-bonding R97:D116	color
receptive	+	+	+	blue
receptive	+	+	−	blue
receptive	+	−	−	blue
receptive	+	−	+	blue
intermediate	−	+	−	green
open	−	+	−	blue
open	−	−	+	blue
open	−	−	−	blue

Table 5.2: Definition of side chain interaction network states involving residue R97 near F pocket floor. A contact criterion of 2.5 $\AA$ between H−bond donors on R97 and acceptor oxygen atoms on an acidic residue was taken as a putative H−bonding contact.

any oxygen atoms of the acidic D114/D116. The opening motion of the F pocket in US-simulations is coupled to a characteristic change in the side chain interaction network: This results first in an intermediate state where R97 is in contact with D114 only and finally in an open switch state with R97 forming a salt bridge with either D114 and D116 simultaneously or only with D116 (see Fig. 5.3 and definition in table 5.2).The distribution of sampled open, receptive and intermediate states correlated with the F pocket width. Thus, we observed that an artificially induced global change of the F pocket provoked a local switch in the basic and acidic side chains at the F pocket. No other significant changes in the average side chain conformations around the F pocket were observed. The induced opening of the F pocket also resulted in a correlated opening of the adjacent C pocket in the majority of sampled conformations. However, in particular at opening distances above 15 $\AA$, there were also states sampled with similar average F pocket size but different C pocket distances (Fig. 5.3: distance between backbone centers of mass of residues 150-156 in the α_2 and 67-73 in the α_1 helix. These types of open structures differ in the angular placement of the α_{2-1} helix segment relative to the α_{2-2} helix segment that flanks the C-pocket (illustrated in Fig. 5.4). Interrupting the continuous helix these ḱink pointsín the long α_2 helix (at positions 151, 152) appear to enhance the types of possible structural alterations in the absence of a bound peptide.

5.4 Accelerated side chain flips of R97 induce a global opening of the F pocket of B*44:02

A statistical correlation between side chain interactions and an induced global shift of the α_2-1 segment does not prove any causal relationship between the two conformational motions. In the previous paragraph we provided evidence that induced opening of the F pocket results in rearrangements of the side chain interactions of R97. To prove a causal relationship it is necessary to show that induced changes in the side chain arrangement can in turn lead to global opening of the F pocket. In order to

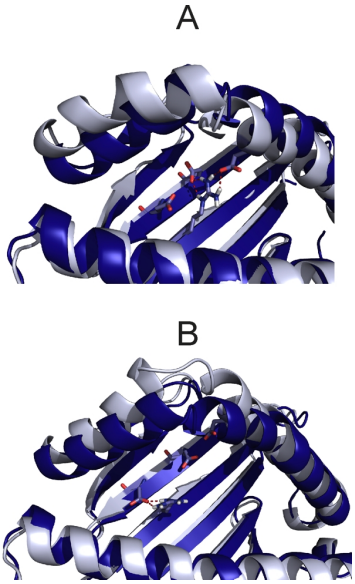


Figure 5.4: Superposition of two types of B*4402 conformations with enlarged F pocket (light blue cartoon) onto the X-ray starting structure (dark-blue cartoon). (A) Example of a conformation with increased distance between $\alpha-$helical segments that flank the F pocket (increased distance between residues 74-85 and 138-149) but a C pocket size similar to the width observed for the X$-$ray structure. This conformation is compatible with the intermediate side chain state of R97 and even with the receptive arrangement (indicated in A). For the more frequently observed conformation (B) with an enlarged C and F pocket, the open side chain state (less frequently intermediate side chain state) was usually found.

further investigate this relationship we performed replica exchange based MD simulations that accelerate side chain changes by the application of a biasing potential in the replica runs. The biasing potential lowers the barrier for side chain dihedral changes around the χ_1 dihedral torsion angle. Different levels of such potential are applied in replica simulations and are allowed to exchange with neighboring replicas including one simulation that runs under the control of the unbiased original force field (only this one is used for the analysis). Due to the exchanges with a reference replica, the sampling of alternative side chain conformations is also enhanced in the reference run (see Methods for details). This technique accelerates the speed of sampling but does not disturb the equilibrium sampling of conformations in the reference simulation [153]. In case of B*44:02 and B*44:05, biasing potentials affected the side chains of R97, D114, D116/Y116 and D156, resulting in enhanced sampling of alternative sterically possible and energetically favorable side chain arrangements. Observed open, receptive and intermediate switch states (as defined in table 5.2) closely resembled states already observed during the US simulations that induce the opening of the F pocket. Interestingly, the sampled switch states correlated with an opening of the F pocket and increased F pocket and C pocket distances in a qualitatively similar manner as observed in case of the US simulations for induced opening of the F pocket (Fig. 5.3). Hence, the switching of R97 side chain contacts with acidic residues in or near the F pocket caused or at least initiated an opening of the F pocket (and in part also of the C pocket) in the BP-REMD simulations. The range of F pocket distances is smaller

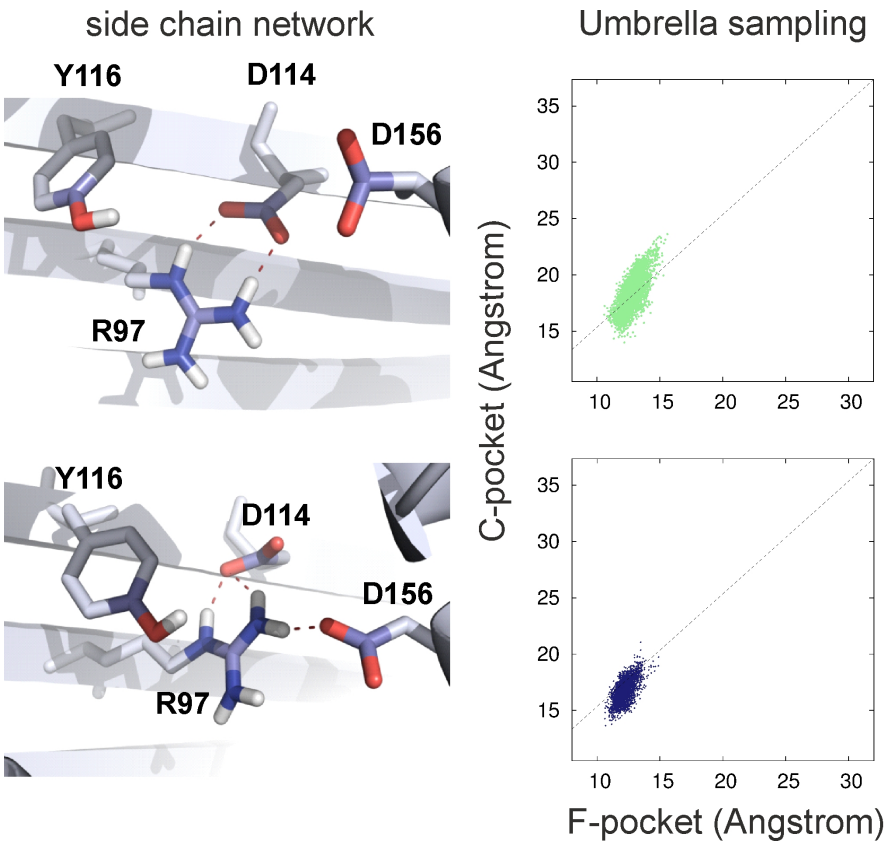


Figure 5.5: F pocket shape and characteristic side chain arrangements of B*44:05 in BP-REMD simulations The first column shows characteristic structural snapshots of intermediate and receptive side chain arrangements of R97 with D114 and D156. The second column (umbrella sampling) indicates the correlation between F/C-pocket size and occurrence of receptive (blue) and intermediate (green) side chain states of B*44:05. The distributions show significant overlap which indicates little correlation of the sampled F pocket size with the side chain configuration.

than those in US simulations of induced global opening. However, this is expected, because distances beyond ca. 20 $\AA$ or below ca. 11 $\AA$ corresponded already to unfavorable states with significantly higher free energy than the F pocket distances around 12-16 $\AA$ and can therefore not be observed in the absence of a stabilizing umbrella potential. The free energy barriers for the allotype B*44:02 in the F pocket distance range of 12-16 $\AA$ are in the order of the mean kinetic energy per degree of freedom (1 RT=0.6 kcal/mol, with R: gas constant and T: temperature of 300 K) and can be overcome by switching salt bridge contacts. In case of BP-REMD simulations of the B*44:05 allotype no significant opening of the F pocket could be induced (Fig. 5.5). The size of the F pocket showed a much smaller deviation from the peptide-loaded shape and only R97 side chain arrangements compatible with the receptive and intermediate switch states, respectively, were observed (Fig. 5.5). The side chain of Y116 restricts the motion of R97 to receptive or intermediate states. Hydrogen bonds to D156 and or D114 dominate, resulting in only slight opening of C and F pockets. This is in line with the US simulations, where receptive and open F pocket states are separated by a larger free energy barrier in case of B*44:05 compared to B*44:02. A close look at

the correlation between F pocket distance and side chain states indicates that even in case of an R97-D156 salt bridge (receptive side chain state) an open F pocket was occasionally sampled due to an alternative opening pathway. This involves a change of the kink angle between the α_{2-1} helix and the rest of the α_2 helix (illustrated in Fig. 5.4).

5.5 Conclusions

Our simulation studies shown here indicate significant differences in the global F pocket flexibility in the absence of a bound peptide between the B*44:02 and B*44:05 class I allotypes. The calculated free energy differences indicated a much smaller barrier for F pocket opening in case of B*44:02 than for B*44:05. Compared to previous simulations on the isolated α_1-α_2 peptide binding platform, we found overall larger free energy barriers for opening which indicate that the presence of the α_3 domain and the $\beta_2 m$ subunit in the complex stabilizes the receptive through a similar free energy contribution for both allotypes. Thus the $\beta_2 m$ subunit assists peptide binding like a chaperone function. This result is also in line with experimental data that indicates a coupling of peptide binding and assembly of the whole MHC class I loading complex (with several components that may act as chaperones, Wright et al., 2004). Further analysis of the molecular mechanism of the global opening of the F pocket in case of the B*44:02 allotype suggested a coupling to changes in salt bridge contacts between R97 and the acidic residues D114, D116, and D156. A receptive state associated with a contact between R97 and D156 dominates a B*44:02 structure with an F pocket close to the bound form (receptive state), whereas a switch to a contact of R97 with D114 and/or D116 resulted in the sampling of mainly open F pocket states (large distances between flanking alpha-helical segments). The causal coupling between global changes in the F pocket size and local salt bridge switches was demonstrated by either inducing a global change (during umbrella sampling) resulting in local changes of the salt bridge network or by accelerating local side chain transitions (using BP-REMD), which in turn leads to global changes of the F pocket. Interestingly, coupled side chain and main chain conformational changes involved in the peptide loading mechanism have recently been observed in the closely related MHC class II molecules (Pos et al., 2012). For human MHC class II molecules (e.g. the human allotypes: HLA-DR1, DR3 or DR4) peptide loading and editing is mediated by the chaperone protein HLA-DM. The crystal structure of a complex between HLA-DM and HLA-DR1 [162] demonstrated the binding of HLA-DM near the P1 pocket of class II. Similar to the F pocket for class I molecules (which anchors the C-terminus of a bound peptide) the P1 pocket in class II is a key specificity pocket that anchors the N-terminal side chain of the bound peptide in class II molecules. The comparison of the HLA-DM/HLA-DR1 complex with isolated HLA-DR1 structures indicates large protein main chain changes and coupled side chain changes in an alpha-segment that flanks the P1-pocket [162]. In particular, the side chain of W43 in HLA-DR1 switches between two rotameric states that are each compatible with only one of the main chain conformations observed for the HLA-DM bound state (without a bound peptide) and the peptide-bound HLA-DR1 conformation, respectively, indicating a causal allosteric coupling. It is likely that coupled side chain and backbone conformational changes play a role in other protein

systems as well. Molecular dynamics simulations offer the possibility to investigate whether statistically associated changes are based on a causal relation or simultaneous changes occur accidentally, without a coupling interaction. Finally, we like to emphasize that in the present case, the coupling mechanism between side chain and global motion is specific for the B*44:02 allotype. However, similar types of coupling may form the basis for allosteric motions in other class I allotypes as well as other systems that involve significant backbone changes upon association and binding of ligands.

Chapter 6

Dipeptides speed up both folding of MHC class I molecules and peptide loading

Information on protein-peptide complexes is of central interest for structural biology and drug development. Peptide binding can even induce the folding of its molecular target: high affinity peptides (consisting of 8-10 amino acids) are known to speed up the folding of MHC class I molecules into their peptide receptive form in vitro. Experiments–performed by Prof. Dr. Sebastian Springer and Ph.D. Sunil Kumar Saini–demonstrated dipeptides to fulfill the minimal requirements to enhance the folding of MHC class I molecules and to promote the peptide exchange. First part of this chapter includes an overview on these experimental findings, which inspired us to perform molecular dynamics (MD) simulations and MMPB/SA free energy calculations. As presented in the second part of this chapter, these simulations could give insight into the molecular mechanism of dipeptide binding.

6.1 Motivation and outline

MHC class I molecules play an important role in the immune system. Located in the endoplasmic reticulum, these proteins consist of a transmembrane α_3 domain, a binding pocket and a not covalently bound $\beta_2 m$ unit and are responsible for the viral and antitumoral immune response. Cell destruction of infected cells is induced by MHC class I molecules that stably bind viral/tumoral peptides into their pocket to transport them to cell surface and present them to cytotoxic T killer cells [62, 39]. Binding peptides have in common a length of 8-10 amino acids as well as characteristic sidechains at key positions that form contacts with certain subpockets of the binding pocket [198, 10, 167]. Natural or artificially designed high affinity peptides proved to induce the folding of denatured bacterial MHC class I molecules in vitro [190, 39]. Thus, high affinity peptides are an effective tool to fold MHC class I molecules into their native structure in vitro. In vivo peptide loading on MHC class I molecules is allotype-dependent and usually assisted by a chaperone protein complex (PCL) that

includes Tapasin (see chapter 5). However, not only high affinity peptides bind to MHC class I molecules: in the presence of the PCL lots of suboptimal peptides with inappropriate length or anchor side chains first bind into the binding pocket, to gradually exchange with high affinity peptides 5) [221, 163]. Thermal denaturation of MHC class I molecules was measured by tryptophan fluorescence (TDTF) experiments to demonstrate the minimal requirements for these suboptimal peptides: a dipeptide glycyl-leucine(GL) successfully promoted the folding of denatured MHC class I molecules into their native form. The label free TDTF method records the fluorescence of the MHC class I molecule solution at different temperatures [30, 46]. The melting (thermal denaturation) temperature of proteins, which is characterized by a dip in fluorescence as soon as tryptophan is exposed to solvent, proved to be coupled to the stability of MHC class I molecules [180]. In a first folding step denatured MHC class I molecules HLA-A*02:01 (A2) were incubated with (1) no peptide, (2) GL, (3) GG and (4) NLVPMVATV in four independent measurements. Afterwards, MHC class I molecules were separated from free dipeptides by mass spectrometry. To stabilize folded MHC class I molecules, the high affinity peptide NLVPMVATV was added to the solution after folding (except for NLVPMVATV) in a second post-folding step. Finally, the fluorescence was recorded for different temperatures. If no peptide or the dipeptide GG is present during the folding step, A2 molecules remain denatured: even the presence of the high affinity peptide NLVPMVATV during the post folding step could not promote the A2 folding (in both cases$\approx$ 15% of A2 molecules are folded.). After the high affinity peptide NLVPMVATV had been added in the first folding step, TDTF measurements showed A2-NLVPMVATV complexes to be stable (melting temperature). Strikingly, the same is true for dipeptide GL: stable native A2-NLVPMVATV complexes (identical melting temperature of $\approx 56°$) formed during the post folding process, after dipeptide GL had been added in the folding step (60% of MHC class I molecules are folded compared to NLVPMVATV). The same experiments were repeated for K^b in complex with FAPGNYPAL and showed identical effects at sightly different melting temperature ($\approx 51°$): the presence of dipeptide GL promoted folding even more successfully than the high affinity peptide NLVPMVATV (110% of MHC class I molecules are folded compared to NLVPMVATV. In contrast, a folding with GG or without peptide resulted in 30% of folded peptides only).

Moreover, fluorescence anisotropy experiments showed the presence of GL to provoke enhanced binding rates of high affinity peptides in vitro. Therefore K^b and A2 allotypes were folded with dipeptides. Kinetic association values (k_{on}) resulted from measuring fluorescence anisotropy of fluorophore labeled high affinity peptides that were added after folding reactions. The presence of GL enhanced NLVPMVATV binding to the A2 molecule and resulted in a kinetic association value $k_{on} \approx 22 * 10^3(M^{-1}s^{-1})$ that was seven fold higher compared to GG (and no peptide) with a kinetic association of $k_{on} \approx 3 * 10^3(M^{-1}s^{-1})$. For the K^b allotype dipeptide GL could speed up peptide binding six fold compared to GG (The same is true if no peptide is present).

GL even proved suitable to prolong the lifetime of functional MHC class I molecules at the surface of murine and human cells and prevent the removal of native complexes from the cell membrane. (MD simulation results described below do not refer to this experimental finding).

All experiments concerning dipeptides and their influence on MHC class I molecules were performed in the Springer lab and inspired us to investigate the role of dipeptides

on MHC class I molecules by means of MD simulations.

6.2 Simulation results and discussion

MD simulations could give insight into the molecular mechanism of dipeptide binding. We followed the hypothesis that the dipeptide GL binds at the same position as two C-terminal amino acids of full length peptides (see Fig. 6.1 A,B,C): acting as a placeholder for the full length peptide, GL keeps the the F pocket in shape via an H-bond network, that resembles the one of the full length peptide. In 20 ns classical MD simulations GL did not leave the binding pocket of K^b or A2. While GG dissociated from K^b within 1.2 ns, this effect was not observed for A2. The width of the binding pocket

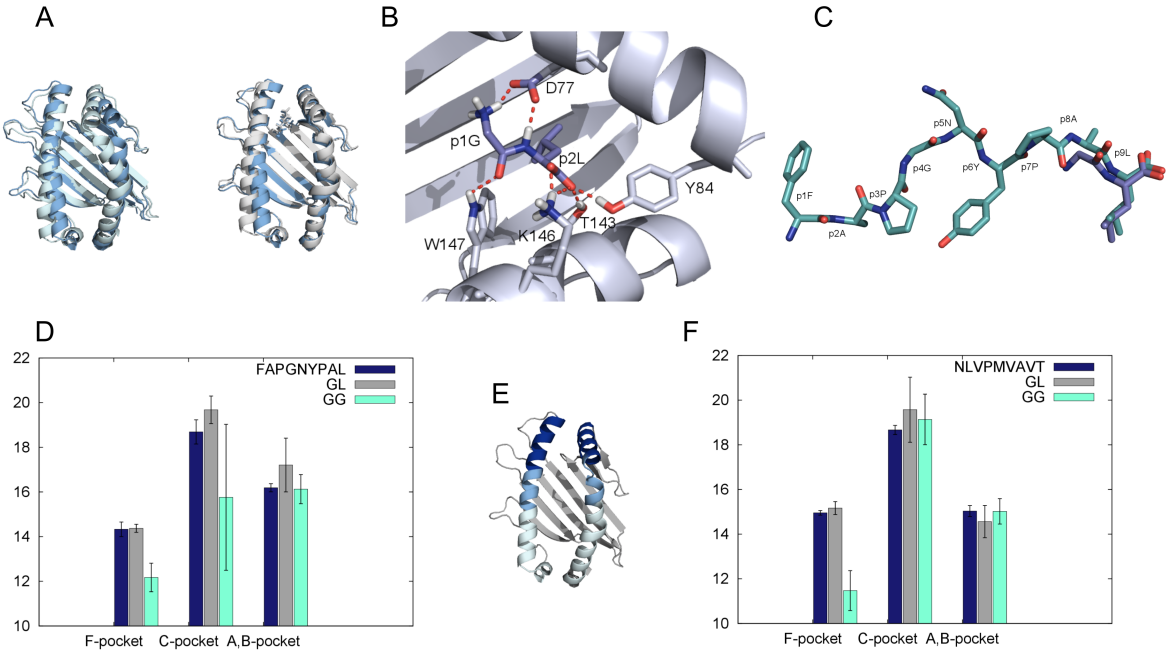


Figure 6.1: MD simulation data:
(A) While GL keeps the binding pocket in shape, GG dissociates after 1.2 ns. Snapshot are taken from the MD simulation with GL (right, gray) and GG (left, light blue) after ∼16 ns. Dark blue binding pocket shows starting structure, which is identical in both simulations.
(B)Proposed binding mode of dipeptides: GL acts as a placeholder for the full-length peptide.
(C) FAPGNYPAL and GL bound to the F pocket(not shown): the position of GL indeed meets the one of the C-terminal amino acids of the full length peptides.
(D,F)(Left) Average F, C, and A,B pocket width of GL, GG and FAPGNYPAL(D)/NLVPMVATV(F) complexed with K^b (D) and A2(F) binding pocket (correspond to distances between the C_α centers of mass of opposing helical segments in the F pocket region (residues 74-85 vs. 138-149; dark blue helices in (E)), the C pocket region (residues 67-73 vs. 150-156; cyan helices in (E), and the A/B pocket region (residues 50-66 vs. 157-174; light blue helices in (E))

allotype	Full length peptide	GL	GG
K^2	-46.08± 7.06	-22.98± 5.0489	-15.04± 4.90
A2	-65.77±6.88	-12.75± 5.04	-2.93±4.50

Table 6.1: Free energy of binding of full length peptide FAPGNYPAL/NLVPMVATV to K^b/A2 as well as dipeptides GL and GG. Resulting form MMPB/SA calculations these values could reproduce the experimentally observed trend.

in the F, C and A,B region (see Fig. 6.1E) depended on the type of bound dipeptide. As shown in Fig. 6.1 D and F, bound GL kept the F pocket in shape. In case of K^b, GL even had a stabilizing effect on the C pocket. Starting structures for the GG and GL MHC class I complexes were created by cutting the full length peptide to AL and changing(= deleting appropriate atoms) A to G and L to G respectively. The HLA-A0201 ILKEPVGHV (protein data bank entry: pdb 3GSO [57]) crystal structure complex serves as starting structure for the simulations of the full length peptide and GG. The GL receptor complex is created by adapting the 2C7U [81] full length peptide complex. MD simulations were performed in explicit water following the following standard protocol: using a parm03 force field [43], each complex was placed in an octahedral TIP3 water box (surrounded by a minimal water layer of 9 Å) and neutralized with sodium (14/15/14) and chloride (6/7/6) ions. After an energy minimization, each complex was positionally restrained (25 kcal mol^{-1} Å^{-2}) and heated from 100 to 300 K in three 0.1-ns simulation runs. Restraints were gradually removed in five MD simulations. Each complex was equilibrated for 1 ns and simulated for 20 ns, whereby particle mesh Ewald method accounted for long-range electrostatic interactions [35]. Binding free energy calculations were performed using the MMPB/SA tool of the amber simulation package [132]. The binding free energy of the GL and FAPGNYPAL K^b complexes was calculated using the 5 ns trajectories. Since for the calculation of the free energy only the bound peptide K^b states are relevant, the free energy was determined evaluating the very beginning of the GG K^b simulation only (0.4 ns). As shown in table 6.1, MMPB/SA calculations could reproduce the trend observed in experiments.
GL speeds up folding and binding of high affinity peptides to MHC class I molecules. We assume GL to bind into the F pocket at a similar position as the full length peptide. Similar H-bond patterns could stabilize the GL-MHC class I complex. Additionally GL enhances binding of high affinity peptides presumably by keeping MHC class I molecules in a peptide receptive shape. Thus, in vitro dipeptides could play a similar role as Tapasin within cells, namely by keeping MHC class I molecules in a peptide receptive form (see chapter 5).

6.3 Acknowledgements

Parts of this chapter have been published in [181]:

Chapter 7

Rapid alchemical free energy calculation employing a generalized Born implicit solvent model

Absolute binding free energy values of diverse peptide-protein complexes can be estimated by post processing molecular explicit solvent simulations using generalized Born (GB) or Poisson Boltzmann (PB) implicit solvent models. In chapter 6 we made use of the MM/PBSA MM/GBSA method to successfully classify binding affinities of peptides of different length for MHC class I molecules. However, the resulting "pseudo" free energy values are approximations of rigorous free energy values only. Inaccuracies are based on simplifications such as rough entropy estimates that enter into the free energy equation. A second source of error is of statistical nature: The averages of absolute free energy values from the complex and its partners sum up to the final binding free energy with propagating errors. To provide more accurate rigorous free energy values, the energy changes associated with alchemical transformations are calculated based on free energy perturbation (FEP) simulations and a GB implicit solvent model. Therefore, Lennard-Jones, Coulomb and Born radii parameters are transformed gradually in a single-topology series. Integrated into a replica exchange scheme, these relative free binding energy calculations converge rapidly.

7.1 Motivation and outline

One of the most important applications of molecular simulations in biophysics is the estimation of free energy changes associated with a structural change in a biomolecule or the association of two binding partners. Often it is of particular interest to predict the effect of a chemical modification of ligand or the substitution of an amino acid residue in a protein on the binding affinity. In alchemical simulation studies it is possible to gradually switch between a force field (Hamiltonian) representing a reference ligand with known binding affinity to a partner molecule and a Hamiltonian representing a

chemically modified ligand. A series of not necessarily physical intermediate states is bridging the conformational gap between initial and final states. These intermediate Hamiltonians are typically described by an associated coupling parameter λ that varies between 0 and 1 and represents at λ=0 and λ=1 the two end-points of the transformation (e.g. modified and unmodified ligand molecule). A variety of well established techniques is available, to extract free energy differences associated with such force field change, including free energy perturbation (FEP) [238], thermodynamic integration (TI) [96], umbrella sampling [209] WHAM [104, 176], and Bennett acceptance ratios (mBAR) [12]. The standard TI method requires the derivative of the force field function with respect to λ at the intermediate stages in order to integrate the free energy change along the reaction coordinate λ [32]. A relative free energy change can be extracted using a thermodynamic cycle that involves an alchemical transformation in the ligand binding site and of the ligand in free solution. It is also possible to obtain absolute binding free energies of complete ligands [38]. The free energy simulation methodology in combination with an explicit solvent representation is widely used to study biomolecular systems and has also found many applications in the drug design field [68, 37]. However, the computational demand is high and more approximate methods have been designed that aim at either reducing the number of intermediate states of a transformation [32, 197] or consider only the end-states of the transformation, e.g. only the force field for the modified and unmodified ligand. The most widely used end-state methods are the Molecular Mechanics Poisson-Boltzmann/Surface area (MM/PBSA) or the Molecular Mechanics Generalized Born/Surface area (MM/GBSA) approaches. MM/PBSA and MM/GBSA make use of snapshots of MD trajectories of molecules in solution [195, 56, 132]. In post processing procedure after removing explicit solvents from each frame the snapshots are evaluated using a continuum solvent description based either on the finite-difference Poisson-Boltzmann (FDPB) or the generalized Born (GB) models and a surface-area dependent nonpolar solvation contribution [68]. The removal of explicit solvent molecules in the post processing yields a large reduction of the number of particles and therefore results in a reduction of the fluctuations of the solvent response to the molecular configuration in each snapshot. For example binding affinities of a ligand to a protein can be estimated from the differences of average energies of the complex and the corresponding averages of the isolated components (over the snapshots of the trajectories). Since this procedure yields only average energies it is often complemented by calculating differences in conformational entropies obtained form a normal mode analysis of the complex and isolated partners. Such approach and related methods like the linear interaction energy (LIE) method require only one MD simulation (or three if separate simulations are performed for complex and partner molecules) but do not yield rigorous free energy changes associated with ligand binding or residue mutation. Another severe drawback is the fact that binding free energies (or free energy changes due to chemical modification or residue substitution) are obtained as small differences between large numbers for the average energies of the complex and separate partners in MM/PBSA and MM/GBSA calculations. Thus even small conformational changes of biomolecules can easily result in high free energy fluctuations [196] provoking large statistical uncertainties and free energy errors that exceed the actual magnitude of the estimated free energy difference. Nevertheless, the MM/PBSA and MM/GBSA methods are widely applied to analyze the stability of protein conformations as well as protein ligand binding affinities [56, 132].

Direct rigorous free energy simulations integrate chemical modifications or residue substitutions by gradually transforming one force field representation into another in the frame of the free energy calculation methods discussed above. While usually coupled with explicit solvent models, such strategies have rarely been combined with an implicit solvent model. Sampling of free energy changes along several intermediate steps should converge rapidly in implicit solvent since the solvent response is instantaneous and does not require an extended averaging over many explicit solvent configurations. Since only the interactions of modified chemical groups are accumulated in rigorous free energy simulations free energy estimates are not associated with large uncertainties that arise while subtracting large total average energies of whole complex and isolated partners in MM/PBSA or MM/GBSA calculations.

However, performing a free energy simulation with the implicit solvent model can be difficult and time consuming if solving the FDPB equation frequently (numerically time consuming) and to extract sufficiently accurate forces from the FDPB model [141, 142]. This is not the case for the GB solvation model that serves as basis for many MD simulation studies but is much less frequently used for free energy simulations. An analytical generalized Born plus non-polar (AGBNP) model provides the basis for estimating absolute ligand binding free energies by gradually decoupling receptor and ligand in a MD simulation using the BEDAM approach [53] or during docking simulations [187, 188].

In the GB-model a chemical group (e.g. single atom) interacts with the system via non-bonded Lennard-Jones and Coulomb potentials. Additionally, each atom is associated with a Born radius, that influences both its solvation and the solvation of other surrounding atoms or groups of atoms. Thus, annihilating a chemical group in a rigorous free energy simulation implies both removing its Coulomb and Lennard-Jones interactions and its associated Born radius. Since the effective Born radii of atoms change in a rather complex functional relation with respect to the coordinates it might be complicated to extract the analytical expression of the derivatives vs. alchemical reaction coordinate λ during MD simulations. Michel et al. [131] circumvented this difficulty by calculating a finite-difference approximation of the derivate vs. λ. Monte Carlo (MC) free energy simulations of ligand binding and predicted binding affinities of a set of inhibitors to cyclooxygenase2, to neuraminidase [131] and cyclin-dependent kinase2 [130] and set of compounds to an estrogen receptor in implicit solvent.
Alternatively an atom without charge and Lennard-Jones interactions could be represented as low dielectric cavity assuming that its influence on calculated free energy differences is small. In the present study the FEP method was combined with a GB implicit solvent model to calculate relative solvation free energies and relative binding free energies. Following an approach by Meng et al. [129] we employed the H-REMD method, that improves sampling [152] and yields the appropriate ensemble averages for the free energy differences per step in λ on the fly as part of the replica exchange criterion [129]. The designed protocol scales charges, Lennard Jones interactions and Born radii simultaneously in different powers of the coupling parameter λ to allow a smooth annihilation or creation of atoms during GB implicit solvent simulations. Relative free energy estimates of hydration of amino acid side chains in tripeptides, relative binding free energies of different FKBP binding ligands and alanine scanning of p53 peptide binding to the oncoprotein MDM2 converged rapidly within less than 1 ns per λ step and were in good agreement with experiment. Combined with a graphical pro-

cessor based hardware replica exchange FEP simulations with a GB continuum model (RE-FEP-GB) are about to allow for calculating relative free energies within minutes of computer time.

7.2 Materials and Methods

7.2.1 Molecular dynamics simulation and simulation systems

Simulations were performed by means of the sander module of the Amber12 package [27]. To calculate relative hydration free energies of amino acid side chains tripeptides with sequence ACE-X-NME were generated using the Amber tleap module. Here the amino acid of interest (X) is flanked by N-terminal Acetyl (ACE) and C-terminal N-Methyl(NME) capping groups, respectively. The short alpha-helical peptide of the protein p53 (PMI-ligand) in complex with the protein MDM2 (pdb3EQS, [157]) served as test system to calculate binding affinities of a series of residue mutations (alanine-scan). Additionally, the RE-FEP-GB approach was tested on a set of small organic ligands in complex with the FKBP binding protein (BUT,4-hydroxy-2-butanone) PDB 1d7j; (DSS, methyl sulphinyl-methyl sulphoxide) PDB 1d7i; (DMSO,dimethylsulphoxide) PDB 1d7h [23]). Simulations were performed using the parm10 force field [31] and the OBC (Onufriev, Bashford and Case) generalized Born implicit solvent model ([149], igb=5 in the Amber input file), wherein vacuum and solvent differ by their dielectric constants ($\epsilon^{solvent} = 78.5$ $\epsilon^{vacuum} = 1.0$). Bonds involving hydrogen atoms were constrained by the Settle algorithm [133] with a time step of $0.001ps$. A Langevin thermostat with a collision frequency of $5.0ps^{-1}$ regulates temperature. Positional restraints were subsequently removed in five consecutive $0.1ns$ simulations. During a final $0.1ns$ equilibration run only C_α atoms were restraint with a force constant of $0.5kcal/\mathring{A}^2mol$, resulting in the input structure for the RE-FEP-GB simulations. While non bonded interactions were calculated up to a cutoff of $99\mathring{A}$, Born radii were considered up to a cutoff of $15\mathring{A}$. The same conditions applied for the RE-FEP-GB simulations.

7.2.2 Free energy perturbation (FEP) in the REMD framework

Relative free energy differences were calculated by exploiting a thermodynamic cycle. In case of relative free energies of hydration it involved creation or annihilation of a chemical group both in aqueous solution (represented by a GB continuum model) and in vacuum. Similar, for relative free energies of binding ligands were modified at a binding site and in free solution. In the FEP scheme free energy changes between two states i, i+1 are defined as

$$\Delta G_{i,i+1} = -k_B T \ln < \exp -(H_{i+1}(x) - H_i(x))/k_b T >_i \tag{7.1}$$

where H is the Hamiltonian in state i and state i+1, respectively. In case of an alchemical transformation these states are appropriate intermediate Hamiltonians on

the path of a complete creation or annihilation of a chemical group in N simulations. It has been recognized by Meng et al. [129] that the FEP expression for the free energy corresponds to the exchange criterion used in H-REMD simulations. Hence, a complete FEP simulation can be setup as an H-REMD simulation with each intermediate represented by one replica. An additional advantage is that the exchanges during H-REMD may improve the convergence of the free energy simulation. The replica exchange FEP simulations with a GB continuum model (RE-FEP-GB) were typically performed with 32 replicas and an equilibration of 0.1 ns followed by a 1.5 ns production run per replica. Exchanges were attempted every 1 ps and accepted according to a Metropolis criterion. Free energy differences in both annihilation and creation direction were extracted. In the H-REMD scheme two free energy differences are calculated for exchange partners $H_{i+1}(x)$ and $H_i(x)$: While energy differences of the conformation i in replica H_i and H_{i+1} enter into $\Delta G_{i\rightarrow i+1,forward(orup)} = -K_bT \ln < \exp -(H_{i+1}(x) - H_i(x)) >_i$, energy difference of the conformation $i+1$ in replica H_{i+1} and H_i determines $\Delta G_{i+1\rightarrow i,backward(ordown)} = K_bT \ln < \exp -(H_{i+1}(x) - H_i(x)) >_{i+1}$. As pointed out in [129] differences between forward and backward free energies are not only incorporated directly in the REMD scheme, but can also serve as a good estimate to check convergence of free energy calculations. The error of calculated free energy changes were also independently estimated by splitting the FEP data into four blocks and calculating block averages and standard deviation.

7.2.3 Transformation potentials in REFEP-GB simulations

For creation/annihilation of atoms only the non-bonded interactions of the chemical group were coupled/decoupled from the system. This involved appropriate scaling of Lennard-Jones parameters, partial charges and changes in Born radii of the group to generate topology files that represent intermediate states of the systems. Test simulations indicated that a linear scaling of the epsilon and sigma values of the Lennard Jones parameters of the modified atoms is an appropriate choice. To avoid numerical instabilities in case of scaled atoms represented with very small sigma values that may still have a partial charge the charge scaling was completed in 28 steps instead of 32 for the annihilation of a group.

More critical is the simultaneous reduction of the Born radii of disappearing atoms. Such reduction is necessary not to end up with dummy atoms that still represent a region of low dielectric constant (although the region could be filled with solvent). A linear decrease of the Born radii coupled with a quadratically decrease of partial charges resulted in smoothly decreasing electrostatic GB solvation and Coulomb interactions of the modified atoms with the rest of the system. In the GB continuum model a surface tension term proportional to the solvent accessible surface area (SASA) accounts for nonpolar solvation. It turned out that it is sufficient to calculate this contribution from the average SASA at the end-states of the alchemical transformation.

7.3 Results and Discussion

7.3.1 Calculation of hydration free energies of amino acid side chains

The free energy of hydration of several amino acid side chains in a tripeptide context (with N- and C-terminal capping groups) was calculated by annihilating the side chain atoms to alanine including the necessary changes of partial charges, Born radii and Lennard-Jones interactions. The main aim was to check how quickly the calculated free energy differences converged for such small test systems. Since experimental free energies of hydration were determined for isolated amino acid side chains an exact agreement was not expected (taken from [222, 194]). Nevertheless, the calculated hydration free energy differences of the most of the selected side chains agreed with experimental values within a range of 2 kcal/mol. Only Leu and Asn hydration free energies deviated more than > 2 kcal/mol. Errors were estimated by comparing the difference of forward and backward FEP free energies and by calculating the variance of block averages of the free energy vs. simulation time. The difference between forward and backward FEP free energies did not exceed 0.34 kcal/mol and was in most cases well below 0.1 kcal/mol after 1.5 ns data gathering time per λ (table 7.1). A similar small error of the calculated free energy changes was estimated from the block

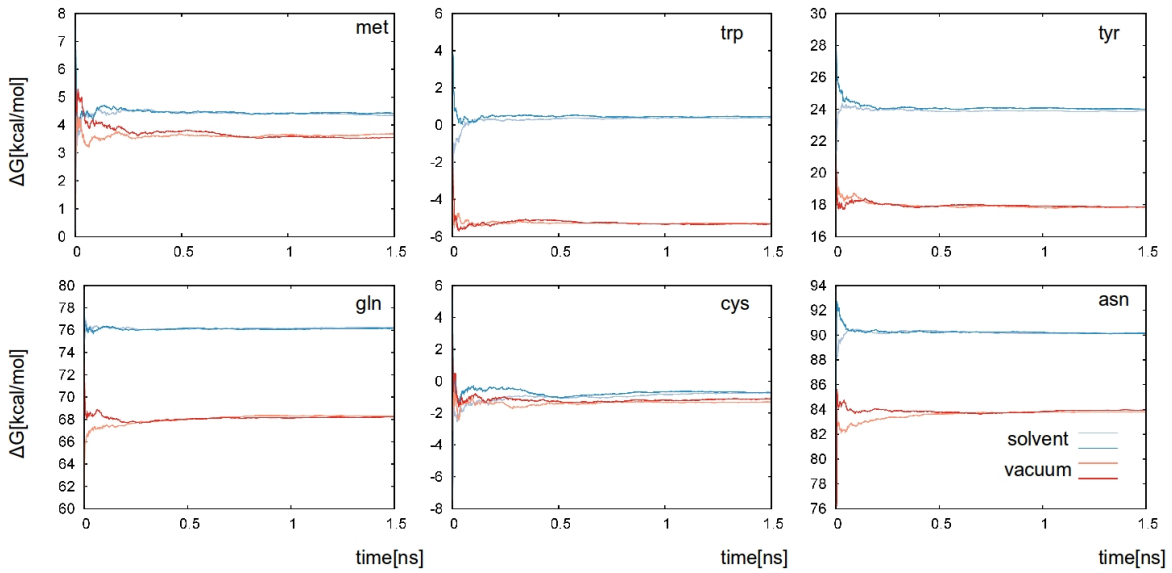


Figure 7.1: Convergence of calculated free energy differences of side chain mutation to alanine in implicit GB solvent and in vacuum. Cumulative forward (side chain annihilation) and backward (side chain creation) free energies in vacuum, $\Delta G^{vacuum}_{up/down}$(reddish), and in GB solvent, $\Delta G^{solvent}_{up/down}$ (dark blue and blue), are plotted versus simulation time. Calculated forward free energies are approaching backward free energies within less than 0.5 ns, indicating rapid convergence. The differences of $\Delta G^{solvent}$ and ΔG^{vacuum} combined with the nonpolar surface area term correspond to the calculated hydration free energy differences $\Delta\Delta G$ in table 7.1.

.

	$\Delta\Delta G$	$error_{up,down}$	$error_{block\ average}$	ΔG_{exp}
leu	0.18	± 0.34	± 0.78	2.28
phe	-0.84	± 0.14	± 0.40	-0.76
met	-0.18	± 0.14	± 0.16	-1.48
cys	-0.27	± 0.20	± 0.29	-1.24
trp	-5.16	± 0.10	± 0.10	-5.88
tyr	-5.58	± 0.16	± 0.15	-6.11
gln	-7.56	± 0.13	± 0.30	-9.38
asn	-5.93	± 0.16	± 0.20	-9.70

Table 7.1: Relative changes in free energy of hydration for mutation to alanine in tripeptide context (in kcal/mol). Experimental data for side chain hydration was taken from reference 30

averaging procedure (table 7.1). As shown in Fig. 7.1, convergence to calculated free energy differences close to the results in table 7.1 was already achieved after 0.5 ns data gathering time (with basically no residual drift in the calculated free energy change). Note, that explicit solvent free energy simulations to calculate hydration free energies of amino acid side chains require much longer simulation times per λ to achieve convergence [187] because it requires sampling over many different explicit solvent configurations.

7.3.2 Relative binding free energies of FKBP ligands

The FKBP protein is a small cis-trans prolyl-isomerase for which crystal structures in complex with the small ligands DSS, BUT and DMSO are available. Both DSS and BUT (see Fig. 7.2) were transformed to the reference ligand DMSO in 32 steps of the coupling parameter λ. The convergence and quality of REFEP-GB free energy estimates depends on the alchemical conversion pathway. In order to get an impression on the changes in the different energy components along the alchemical pathway single

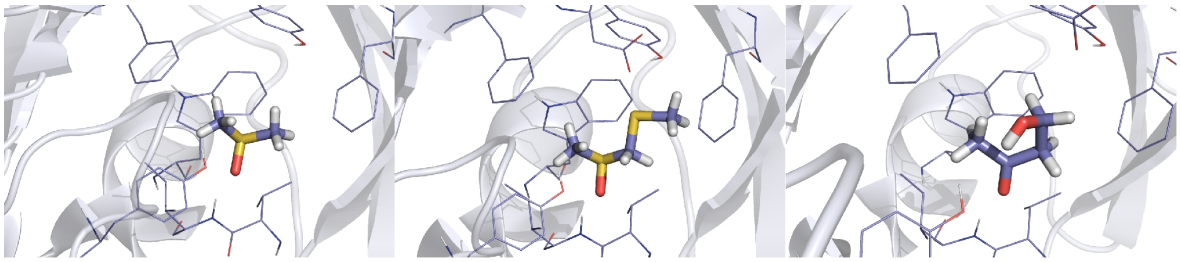


Figure 7.2: FKBP protein in complex with small ligands DMSO(dimethylsulphoxide, left panel), DSS (methyl sulphinyl-methyl sulphoxide, middle panel) and BUT (4-hydroxy-2-butanone, right panel). Ligands are indicated as atom-color-coded stick models.

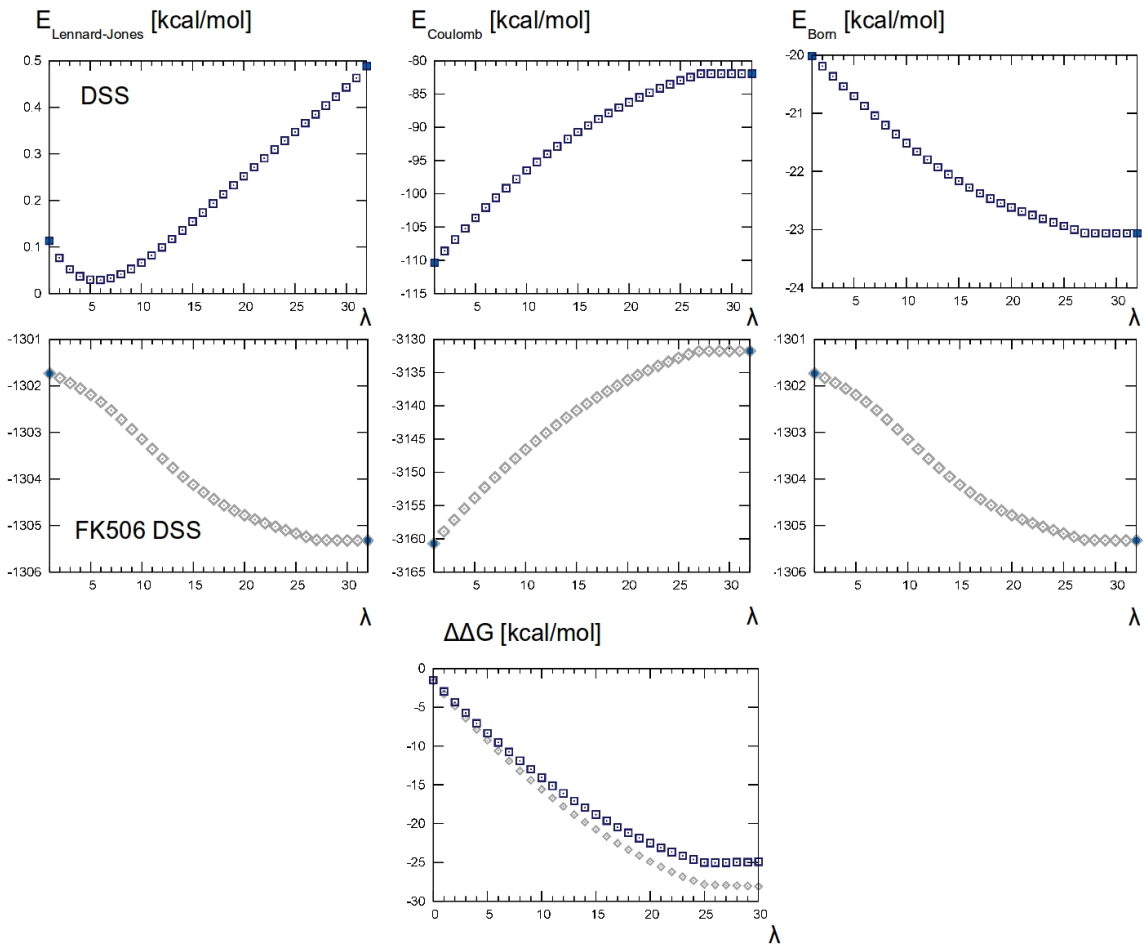


Figure 7.3: Calculated energy contributions for a single snapshot of the alchemical DSS to DMSO transformation along the coupling parameter λ (changing in 32 discrete steps). Energy contributions correspond to Lennard-Jones E_{VDW}, Coulomb E_C and GB reaction field E_{GB} terms and are given for the DSS to DMSO conversion in the absence of FKBP protein (upper row of plots) and same for the conversion in complex with FKBP (second row of plots). The filled bold squares and diamonds at the end points of the transformation indicate calculated energies obtained with the original topology files representing the end states of the alchemical transformation. The calculated change in free energy obtained with the RE-FEP-GB approach vs. coupling parameter λ is indicated in the lower panel (square indicate the conversion in the absence of FKBP protein and diamonds represent the transformation in complex with the protein partner).

conformers for the DSS - DMSO conversion in solution and in complex with FKBP were analyzed. As illustrated in Fig. 7.3 the different energy components along the alchemical transformation pathway change smoothly and the components that make up the largest contributions change almost linearly. Such smooth change resulted in similar free energy changes per step of λ along the whole alchemical path and in rapid convergence along the path. Note, that the Born Radii of disappearing atoms

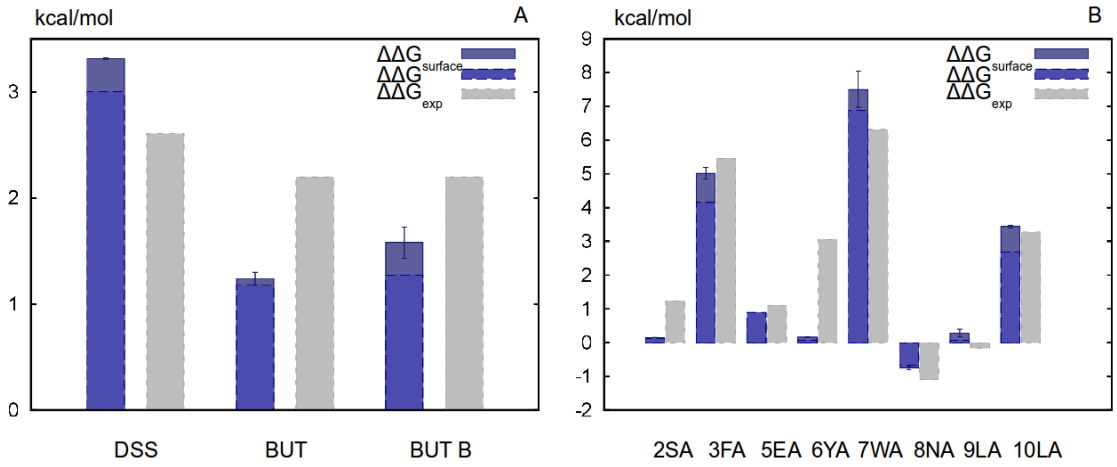


Figure 7.4: (A) Comparison of calculated RE-FEP-GB binding free energy changes (blue bars) for alchemical transformation of DSS and BUT ligands to DMSO in complex with the FKBP protein and experiment (grey bars). Simulations for the BUT system were started from two possible initial orientations of the methoxy group of BUT in the binding pocket. (B) Comparison of calculated RE-FEP-GB binding free energy changes for in silico mutation of PMI peptide side chains to alanine in complex with the MDM2 protein (blue bars) and experimental alanine scanning on the same system (grey bars). Experimental values are taken from [113]. The light blue part of the blue bars indicate the nonpolar solvent accessible surface area contribution $\Delta\Delta G_{surf}$.

were scaled down to 0.1$\AA$ † within 32 replicas. This was necessary to avoid numerical instabilities (the residual energy due to this small Born radius of a dummy atom to the system was in all cases $<$ 0.01 kcal/mol and was neglected). The calculated binding free energy differences due to the alchemical transitions agreed with experimental data within a range of 0.9 kcal/mol. Very similar free energy differences were obtained for the BUT-to-DMSO case starting from two different internal rotameric states of the methoxy group (see Fig. 7.4).

7.3.3 In silico alanine scanning of PMI ligand in complex with MDM2

The α-helical peptide with sequence "TSFAEYWNLL" (in one-letter code, termed PMI ligand) binds specifically to the MDM2 protein. The structure of the PMI/MDM2 complex has been determined experimentally [113] (see Fig. 7.5). Moreover, the stability of the PMI/MDM2 complex and the effect of several substitutions of PMI interface residues on binding affinity has been quantified experimentally‡in an alanine scan [112]. Residues Phe3 (F3) Trp7(W7) Tyr6(Y6) and Leu10(L10) were found to contribute most critically to effective binding. Relative binding free energies were calculated by alchemical RE-FEP-GP transformations of the peptide in complex with MDM2 and in

†slightly above the internal amber offset of $b_{offset} = 0.09\AA$

‡with the SPR based K_d measurements [113]

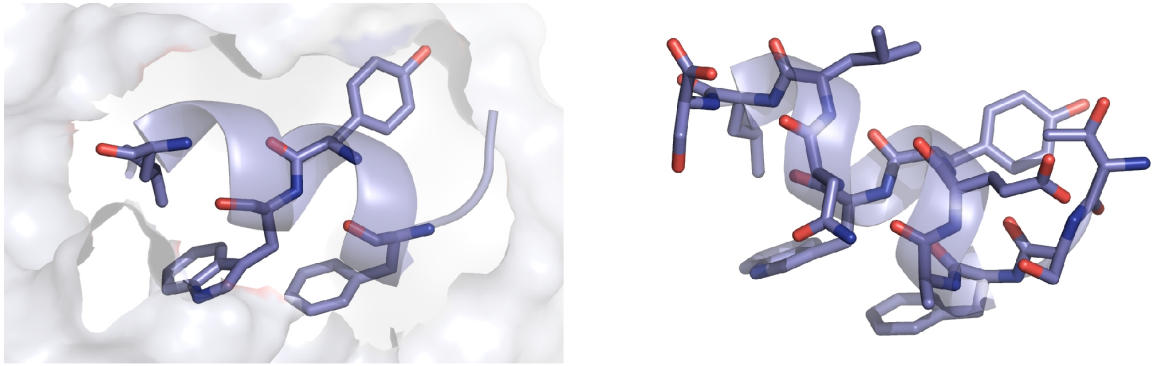


Figure 7.5: Oncoprotein MDM2 in complex with the PMI ligand (left panel). Peptide binding site of the protein MDM2 (grey surface representation) in complex with the α helical PMI peptide (cartoon and stick representation of the most important binding residues). (right panel) Stick representation of all side chains of the PMI ligand.

the absence of MDM2. The calculated alanine scanning free energy changes differed from experiment less than 0.5 kcal/mol except for 3 cases, S2A, Y6A and W7A, with deviations exceeding 1.5 kcal/mol with respect to experiment (Fig. 7.4). Again, calculated free energy differences converged rapidly, characterized by residual differences between forward and backward FEP results (table 7.2 and Fig. 7.6).

For this system the effect of reducing the Born radii during the FEP calculations was compared to calculations in which the final dummy atoms still represent a region of low dielectric constant (no scaling of Born radii) but without a charge or Lennard-Jones interaction. Depending on the type and location of the side chain at the interface and location in the unbound ligand the difference in calculated free energy change for the two cases reached between a few tenth to up to 2 kcal/mol (table 7.2). Also, better agreement with experiment was observed for the calculations including scaling of Born radii.

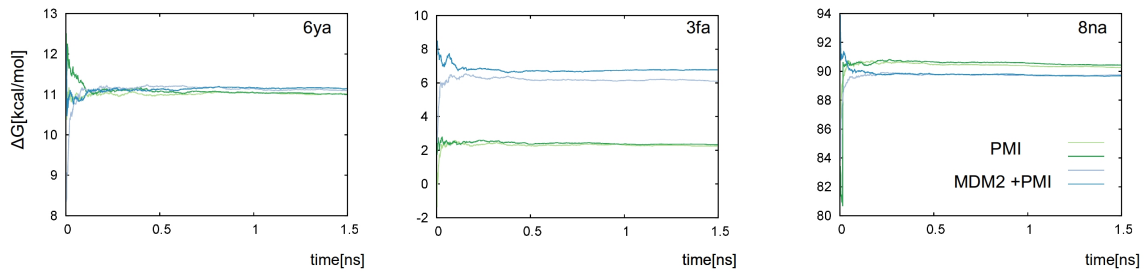


Figure 7.6: Convergence of calculated binding free energy differences for alanine scanning of side chains of the PMI ligand and PMI ligand in complex with MDM2. Cumulative forward and backward free energy changes of three alchemical side chain to alanine conversions (in complex with MDM2 protein: light and dark blue), $\Delta G^{PMI}_{up/down}$, and in solvent, $\Delta G^{solvent}_{up/down}$ (light and dark green), are plotted versus simulation time.

	$\Delta\Delta G_{el,RB(i)}$	$error_{up,down}$	$error_{block\ average}$	$\Delta\Delta G_{el,RB=const.}$	$\Delta\Delta G^{bind}_{exp}$
2SA	0.14	± 0.05	± 0.11	0.17	1.24
3FA	5.11	± 0.63	± 0.24	3.23	5.46
5EA	0.78	± 0.13	± 0.29	-1.17	1.10
6YA	0.48	± 0.55	± 0.53	-1.43	3.06
7WA	7.85	± 0.36	± 0.26	9.48	6.31
8NA	-0.60	± 0.15	± 0.25	-1.66	-1.10
9LA	0.10	± 0.32	± 0.27	-1.54	-0.17
10LA	3.34	± 0.59	± 0.28	2.52	3.28

Table 7.2: Calculated relative changes in binding free energy of PMI peptide to MDM2 protein upon mutation to alanine (in kcal/mol). For comparison free energy differences without changing Born radii of disappearing atoms (RB=const.) are also given. Experimental data was taken from reference 33

7.4 Conclusion

The generalized Born model is usually employed to estimate free energy differences within the MM/GBSA approach, that post processes trajectories. This method typically yields only mean energy differences between end-states of representing for example the binding of two different ligands to a receptor molecule. In the present study we used the free energy perturbation approach in combination with Hamiltonian replica exchanges between intermediate states to evaluate the free energy change associated with an alchemical transformation: Therefore we change partial charges Lennard-Jones interactions and Born radii of atoms that undergo alchemical changes. To achieve highest efficiency it is best to scale all parameters simultaneously in one FEP process. Decreasing Born radii can increase the electrostatic reaction field (at constant partial charge of an atom) and provoke an increase of the electrostatic solvation of a disappearing atom. We circumvent these effects by scaling the charge quadratically combined with a linear change in Born radii to achieve a smooth free energy change along a coupling parameter of the alchemical transformation. Although the GB model accounts for solvation effects only approximately we could demonstrate for several test cases that the approach yields free energy changes in close agreement with available experimental data. Moreover, calculations converged rapidly in relatively short simulation times per step in coupling parameter λ. This is possible presumably due the benefits of the implicit solvent model, that allows for a more rapid sampling of alternative states and for a instantaneous solvent response that does not require averaging over many solvent degrees of freedom during sampling. The present method yields rigorous free energy changes associated with an alchemical change in the system which is a decisive advantage compared to MM/GBSA and related approaches. Accuracy of calculations could possibly increase by combining the present pure GB implicit solvent model with an explicit representation of just a few specifically bound water molecules at the ligand binding site. In addition, it may also be possible to reduce the number of steps in the alchemical coupling parameter λ to further decrease the computational demand. However, simulations in the GB continuum model are already quite rapid

especially on graphical processing units (GPUs) reaching up to 400-500 ns per day for a small to medium sized protein on a single GPU (using the latest Amber-GPU version of the pmemd module). The present REFEP-GB simulations converge within 1 ns per λ step. Hence, a cluster of GPUs would allow performing a single REFEP-GP free energy calculation in 2 minutes simulation time. This puts systematic rigorous free energy applications to test hundreds or even thousands of ligand modifications at reach within a day of GPU computer time.

Chapter 8

Dynamic self-rearrangement of molecules in Molecular Dynamics simulations to identify best binder

The RE-FEP-GB approach proves as an effective tool to estimate relative free energy differences of ligands, that differ by several atoms. As a drawback solute solvent interactions are estimated by an implicit GB model, which both disrespects the role of key water molecules and may fail predicting polar and buried charges. Thus, in this chapter a method is developed to "grow" realistic ligands within their binding pockets in native-like environment respecting explicit water molecules. During the simulations scaffold and molecular groups can combine into realistic molecules, that can degrade and (re-)arrange dynamically interacting with explicit water molecules. The approach aims at distinguishing binding affinity of ligands, that differ by small molecular groups, sticking to the same molecular scaffold. If the scaffold is softly restrained to its experimentally resolved binding pose, small molecular groups are expected to adopt native binding poses. Their binding periods may correlate with their binding free energy. In the following limits and potentials of this approach will be discussed.

8.1 Motivation and outline

Identifying binders for known molecular targets is the central aim of lead discovery [168, 25]: Binding affinities of ligands were successfully improved by fragment based approaches, that rely on the following puzzle-like principle: Both experimental(NMR [226] and X-ray [41]) and computational methods help identifying the binding sites of a number of small low affinity binders. Supposed to bind into sub-pockets of the desired target region, these three dimensional complexes serve as scaffold for the high affinity ligand: Chemical linkage and assembly of these small sub-pocket binders can result in a high affinity ligand for the desired pocket. Inspired by this fragment based approach we aimed to mimic a dynamic self (re-)arrangement of molecule fragments

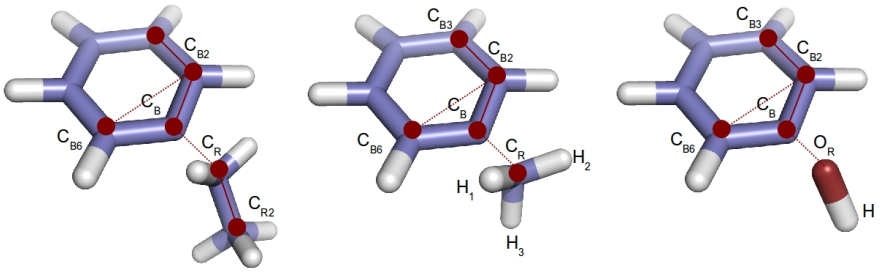


Figure 8.1: Geometrical orientation of ring torso and fragments. During the simulations the methyl, ethyl and hydroxyl (from the left to the right) fragment are free to exchange with the ring torso. Native covalent bonds and mean chemical orientation between C_B and C_R (dashed lines) are replaced by short range potentials during simulations.

in MD simulations. Surrounded by explicit water, these fragments can combine into realistic molecules, that can degrade and (re-)arrange dynamically during a simulation. The presence of a molecular target should principally not only result in a native-like complex: Lifetime of ligands in the desired binding pocket should also correspond to their binding affinities, and thus ideally reflect the free energy scale (i.e. survival of high affinity ligands). Crystal structures of T4 lysozyme in complex with toluene (PDB:3GUK [117]), ethyl-benzene (PDB:3GUL [117]) and phenol (PDB:3GUO [117]) as well as experimental data on its binding kinetics makes its hydrophobic core to a suited molecular target, to prove the principle. Fragments are created by splitting neutral binders into a positively charged benzene ring torso and negatively charged ethyl, methyl and hydroxyl groups, respectively (see Fig. 8.1). (A detailed description of ligand preparation follows in section 8.3). In the following, interactions between the ring and its fragments are driven by short range potentials that mimic geometrical binding properties as well as intermolecular interactions. Advantages and limits of this approach will be discussed basing on the following two test cases: An ethyl, methyl and hydroxyl group are competing to form a complex with the ring torso[†] in solution: Native-like toluene, ethyl-benzene and phenol rings formed in MD-simulations. The same was observed in the the hydrophobic cavity of the binding-pocket with the ring scaffold at its appropriate place: The T4 lysozyme ethyl-benzene and phenol complexes resembled native like structures.

8.2 Short range interactions drive dynamic self (re-)arrangement of molecules

MD simulations are a powerful tool to study the dynamics of biomolecules. Thereby the motion of biomolecules is driven by inter- and intra-molecular forces: While molecular geometry is restricted by harmonic bond and angle potentials and a trigonometric torsion potential series, van der Waals and electrostatic potentials account for intramolecular interactions by preventing molecular overlap and providing for a native-like charge distribution. Solving Newton's equation of motions, positions and velocities can be predicted on the atom scale. While global motions are governed by both dihedral

[†]Benzene ring torso forms by truncating one hydrogen of classical Benzene.

angle changes (several minima the torsion angle potential can result in global protein motion) and intra-molecular interactions, bond lengths and bond angles of neighboring atoms barely fluctuate at physiological conditions. In contrast to quantum-chemical calculations, the chemical structure is kept fixed during simulation time. Restricted by sharp harmonic potentials, covalent bonds do never form or break during MD simulations. We try to break this paradigm by enabling dynamic self arrangement of molecules during a simulation run. Short range potentials control the orientation and the geometrical structure of dynamically combining molecules (i.e. ring torso and fragments). Acting on the acceptor carbon of the ring torso and the receptor carbons (oxygen, respectively) of its fragments, the binding potential V_W is mediating the bond length in MD simulations.

$$\begin{aligned} &V_W(d_ij) = -k_b((d_{ij} - d_0)^2 - (\frac{\Delta_B}{2})^2)^2 \text{ if } |d_{ij} - \Delta_B| > 0 \\ &0 \text{ otherwise} \\ &k_b = \frac{k_{b0}}{(0.5\Delta_B)^4} \end{aligned} \tag{8.1}$$

with force constant k_b, potential width Δ_B (where V_B<0) and native bond length d_0. Centered around the chemical bond length, these narrow Gaussian-like potential wells restrict receptor atoms on a shell around receptor carbon of the ring torso at chemical bond length distance (see Fig. 8.2(A)). Attractive interactions have any effect only if binding fragments are in close proximity ($\pm$ 0.5$\mathring{A}$) of each other. The depth of the potential well is adjusted such that thermal fluctuations still can neutralize its effect. These two aspects are decisive for successful molecular (re-)arrangement. Fragment concentration is not controlled by the number of fragments–only one of each kind is in solution–but by additional potential walls that force the fragments into the vicinity of the ring torso carbon of interest: Additionally this potential prevents charged fragments from getting trapped by charged side chains inside the T4 cavity: Thus, the influence of the artificial additional charges on the ring torso fragment assembly is balanced.

$$\begin{aligned} &V_P(d(\vec{r_{C_s}}, \vec{r_F}) = 2.0(d(\vec{r_{C_s}}, \vec{r_F}) - (d_0 + \Delta_{wall}))^2 \text{ if } 0 < d(\vec{r_{C_s}}, \vec{r_F}) < d_0 \\ &0 \text{ if } d(\vec{r_{C_s}}, \vec{r_F}) < \Delta_{wall} + d_0 \\ &2.0(d(\vec{r_{C_s}}, \vec{r_F}) - d_0)^2 \text{ else} \\ &\text{with } \Delta_{wall} = 3.5,\ 1.5 \end{aligned} \tag{8.2}$$

The resulting total potential is substituting static covalent bonds in MD simulations(see Fig. 8.2(A)):

$$V_B = V_P(d(\vec{r_{C_s}}, \vec{r_F}) + V_W(d(\vec{r_{C_s}}, \vec{r_F})) \tag{8.3}$$

The width of these potential walls differs in the binding cavity and in solution: While fragments exchange effectively within the binding cavity at a potential width of 3.5 $\mathring{A}$, potential walls have to be limited to a width 1.5 $\mathring{A}$ in case of the ring torso fragment in solution. This stronger restriction in solution is necessary to balance the influence of the dipole moments of fragments. A distance dependent angle potential provides for correct chemical orientation(see Fig. 8.2(B)): the first term is of the same shape as the bond length potential and accounts for correct angle formation between three

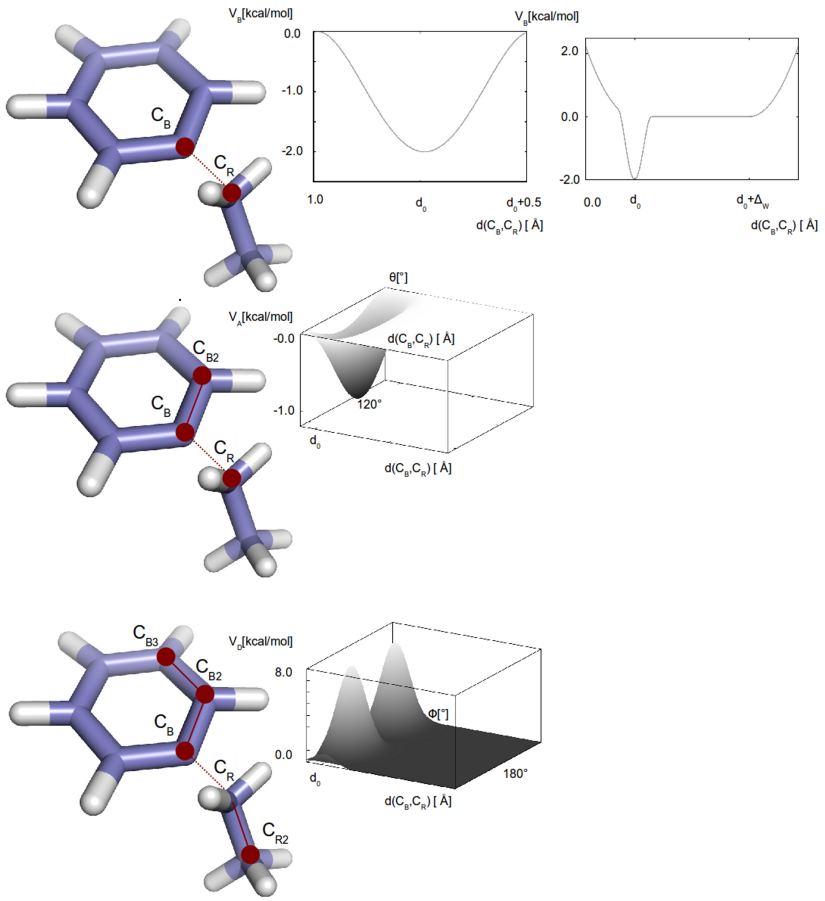


Figure 8.2: (A) Binding potential V_W: Ring torso carbon C_B and fragment torso C_R (O_R) interact via a Gaussian like potential well V_W , that is catching fragments from solution. Binding potential V_B.: potential walls regulate concentrations. (B) Angle potential V_A: Gaussian like angle potential is modulated by a distant dependent Gaussian. It affects the geometry only, if the distance between ring torso carbon C_B and fragment torso C_R (O_R) d(C_B,C_R) equals approximately the experimental bond length(colored dark grey). (C) Dihedral potential V_D: a classical MD potential V_D acts on the ring torso carbon C_B and the fragment torso C_R (O_R) if their distance d(C_B,C_R) is in their experimental range.

neighboring covalently bound atoms. The distance dependent second term eliminates long range interactions: the angle potential is acting only, if the distance between acceptor and receptor atoms equals approximately the bond length ($\pm 0.2\mathring{A}$).

$$V_A(\theta_{ijk}, d_{ij}) = -k_a((\theta_{ijk} - \theta_0))^2 - (\frac{\Delta_A}{2})^2)^2 \exp(-100.0(d_{ij} - d_0)^2) \text{ if } |\theta_{ijk} - \Delta_A| > 0$$

$$V_A(\theta_{ijk}, d_{jk}) = -k_a((\theta_{ijk} - \theta_0))^2 - (\frac{\Delta_A}{2})^2)^2 \exp(-100.0(d_{jk} - d_0)^2) \text{ if } |\theta_{ijk} - \Delta_A| > 0$$

0 otherwise

$$k_b = \frac{k_{a0}}{(0.5\Delta_A)^4} \tag{8.4}$$

with angle $\hat{\mathrm{I}}_{\measuredangle ijk}$, force constant k_b, potential width Δ_b (region, where V_b is acting i.e. $V_b < 0$), native angle θ_0, native bond length d_0 and actual current $C_B C_R$ distance d_{ij}. While harmonic angle and bond MD potentials are replaced by (distance-dependent) short range potential wells, the MD dihedral potential is modulated by a distance-dependent Gaussian, to eliminate long range interactions (see Fig. 8.2(C)):

$$V_D(\Psi_{ijk}, d_{jk}) = -k_\psi(1 + \cos(\eta\Psi + \delta)) * \exp(-100.0(d_{jk} - d_0)^2) \tag{8.5}$$

with force constant k_ψ, phase δ, frequency η and native bond length d_0 . A modification of relevant intermolecular interactions is a necessary prerequisite for effective molecular recombination. To avoid molecular repulsion between the ring torso and the fragments during MD simulations, relevant electrostatic and van der Waals interactions have to be modified[‡]: Each atom of each fragment is excluded to interact with both the acceptor carbon and with its two neighboring carbons (C_{B2}, C_{B6} in Fig. 8.1) of the ring torso. However, atomic interactions between the T4 lysozyme and the ring torso and between the T4 lysozyme and the fragments are fully considered. Thus, recombined molecules can interact classically with binding cavity atoms. To prevent simultaneous binding of different fragments to the ring torso, the interactions of different fragments with each other are fully respected. The acceptor carbon is fishing the carbon (oxygen) atom of one of the fragments out of the solution, keeping it approximately at correct bond length. Only then the angle and dihedral angle potentials start pushing the complex in correct shape.

8.3 METHODS

8.3.1 FRAGMENTS AND RING TORSO PREPARATION

Both the ring "torso" (benzene without 1 hydrogen) and the fragments are constructed respecting the molecular geometry of ethyl, methyl and hydroxyl groups (fragments), respectively. Both molecular mechanics calculations (antechamber) and manual changes approximate their chemical properties only. Electrostatic properties at the artificial interface between ring and fragments are estimated at the expense of exact quantum mechanical calculations. By removing an hydrogen of a benzene ring, positively charged C_B remains unshielded (+0.2e). The charge distribution of the hydroxyl (ethyl-)group was created by imitating the one of a tyrosine (valine) amino acid. Charge distribution around C_β of phenylalanine provides the basis to estimate the methyl group charges. Resulting in a total charge of -0.2e, fragments are neutralized during complex formation with the ring torso and thus resemble their native models. While agreeing in total net charge, atoms of each fragment differ in charge value. The resulting different dipole moments respect different chemical properties of hydroxyl, ethyl and methyl groups.

8.3.2 PROTOCOL

Simulations are performed by means of the amber 12 package (sander) in a ff03.r1 forcefield[43]. Dynamic self (re-)arrangement of molecules is based on the following protocol:

[‡]by manipulating the excluded list in the topology file

1. Topology file
   To respect the non classical geometry and interactions of the ring torso and the fragments, the topology file has to be adapted. Interacting via Lennard-Jones or Coulomb forces, the ring torso C_B and the fragment C_R O_R atoms would overlap and repel each other strongly. To avoid this, the excluded list in the topology files was modified such that these atoms exert intramolecular-like forces on each other. (Note that in the sander 12 MD program the excluded list is updated internally, restoring the original excluded list and disrespecting external changes. Use the $ewtyp = 0$ option to change this.).

2. Minimization of ring torso and fragments in implicit solvent
   The fragments and the ring torso are placed randomly in coordinate space. During an implicit solvent minimization (20000 steps, nb=20, igb=2, cut=900.0, nsnb=20,imin=1, ntmin=2,) fragments are pulled in close vicinity of the ring torso inside a sphere of $7\mathring{A}$, resulting in the starting structure for solvation in explicit water and the equilibration run.

3. Solvation in explicit water and equilibration run
   Thus, ring torso and three fragments are centered in an octahedral box and dissolved in TIP3P water [84], calculating long range electrostatic interactions by means of the particle mesh Ewald option. After an energy minimization the ring and the fragments were heated in 100 K steps (each for 0.05 ns) imposing positionally restrains on all heavy atoms (50 kcal/mol $\mathring{A}^{-2}$). At 300K restraints were removed gradually during additional 0.1 ns simulation time. However, periodic boundary conditions turned out to be explosive (sander bomb) for the production run. We suspect interactions between the fragments and the ring torso to be correctly switched off in short range only. As soon as the ring torso approaches the boundary, the fragments start being re-exposed to actually excluded Lennard-Jones and electrostatic interactions. Thus, the periodicity is broken by the cap option in favor of a stable production run.

4. Cap option for stable production run
   Centered around the C_B carbon of the ring torso, a spherical water cap is kept in shape by a cap restraining potential (50 kcal/mol $\mathring{A}^{-2}$) and thus replacing the continuous solvent. The ring torso is fixed with very weak positional restraints (0.5 kcal/mol $\mathring{A}^{-2}$), to avoid the three fragments to displace the ring torso by exerting forces form different directions. The fragment concentration is not controlled by the number of fragments–only one of each kind is in solution–but by an additional potential that forces the fragments into the vicinity of the ring torso carbon of interest.

8.4 Results and discussion

(Re-)arrangement of the ring torso and the fragments was observed both in solution and in the T4 lysozyme binding cavity. A correct chemical structure arises on average only. While ,on the one hand, guaranteeing for perfect ring torso fragment assembly, steep deep potential wells would disable effective fragment exchanges on the other

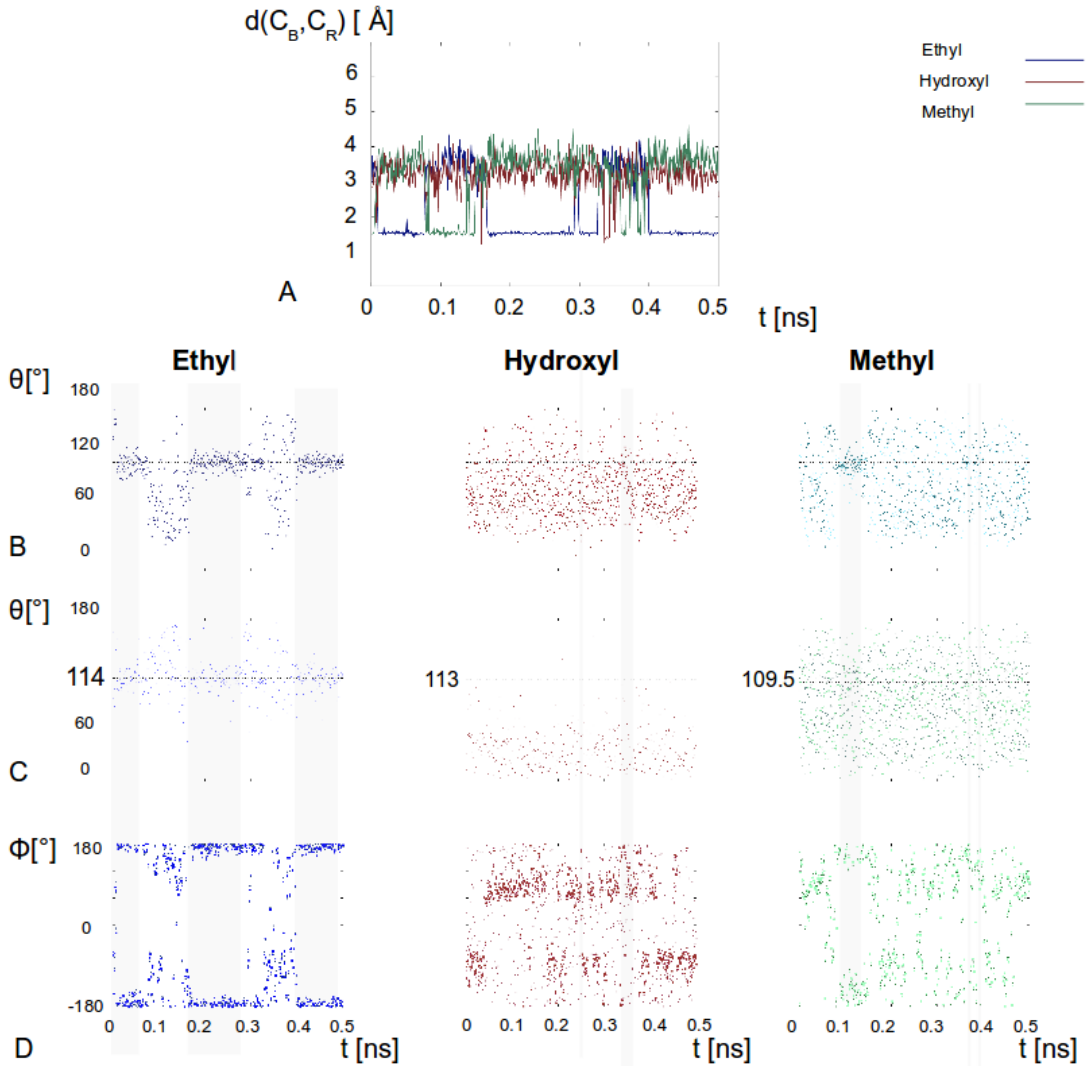


Figure 8.3: Ring torso with one methyl, ethyl and hydroxyl group in aqueous solution.(A) C_B,C_R distance versus simulation time (B) θ[°] 120° angles (C) θ[°] 114° 113° 109.5° angles: In case of hydroxyl and methyl these angles contain hydrogen(D) dihedral angles

hand. Adjusting the depth of angle and bond potential wells in the k_bT range (see table 8.1 8.2 8.3) is essential to keep the balance between effective fragment exchanges and correct chemical structure formation.

8.4.1 Toluene, Ethyl-benzene and Phenol rings

First a simulation of the ring torso with one copy of each fragment type will be analyzed (ring torso +1 ethyl +1 methyl + 1 hydroxyl). In solution the simulation data of fragments and ring torso result on average in native-like toluene, ethyl-benzene and phenol rings (see Fig. 8.3): ethyl-benzene complexes agree on average with the native values in table 8.1 8.2 8.3. As soon as the carbon carbon distance is in the desired range of around 1.52, both relevant angles and torsion angles will adopt desired values (see table 8.1 8.2 8.3 and Fig. 8.3). The same is true for dihedral angles and 120°

Bonds				Angle				Dihedral			
	$d_0[Å]$	k_b	Δb[Å]		$\theta_0[°]$	k_θ	$\Delta\theta\circ$		k_ψ	η	$\delta[°]$
C_BC_R	1.52	0.8/2.0	1.0	$C_{B2}C_BC_R$	120	2.0	60	$C_{B3}C_{B2}C_BC_R$	3.625	2.0	180
				$C_{B6}C_{B2}C_BC_R$	120	2.0	60	$C_{B6}C_{B2}C_BC_R$	1.1	2.0	180
				$C_BC_RC_{R2}$	114	2.0	60				

Table 8.1: Ethyl

Bonds				Angle				Dihedral			
	$d_0[Å]$	k_b	Δb[Å]		$\theta_0[°]$	k_θ	$\Delta\theta\circ$		k_ψ	η	$\delta[°]$
C_BC_R	1.5	0.8	1.0	$C_{B2}C_BC_R$	120	2.0	60	$C_{B3}C_{B2}C_BC_R$	0.16	3.0	0
				$C_{B6}C_BC_R$	120	2.0	60	$C_{B6}C_{B2}C_BC_R$	1.10	2.0	0
				$C_BC_RC_{H1}$ $C_BC_RC_{H2}$ $C_BC_RC_{H3}$	109.5	0.7	60				

Table 8.2: Methyl

Bonds				Angle				Dihedral			
	$d_0[Å]$	k_b	Δb[Å]		$\theta_0[°]$	k_θ	$\Delta\theta\circ$		k_ψ	η	$\delta[°]$
C_BC_R	1.37	0.8 1.0		$C_{B2}C_BC_R$	120	2.0	60	$C_{B3}C_{B2}C_BC_R$	3.625	2.0	180
				$C_{B6}C_BC_R$	120	2.0	60	$C_{B6}C_{B2}C_BC_R$	1.1	2.0	180
				C_BC_RH	113	2.0	60	$C_{B2}C_BC_RH$	2.3	2.0	180

Table 8.3: Hydroxyl

angles, spanned by the ring torso and the hydroxyl and methyl group. In contrast, this tendency is not pronounced in case of 109.5° and 113° angles, that are spanned by a carbon carbon (oxygen) and a hydrogen atom. Its low mass turns a hydrogen atom into the ideal target to be excited by thermal fluctuations. This enhanced mobility could be reduced either by lowering hydrogen angle potentials or by artificially increasing the hydrogen mass. In the first case the influence of the angle potentials on the total energy landscape would be overestimated presumably decreasing the exchange rates of fragments. In the second case an artificially increased mass would provoke unrealistic kinetics. Increasing the mass of relevant hydrogens artificially in some test simulation runs indeed resulted in more precise adjustment of hydrogen involving angles (data not shown). As the position of the fragment torso in the binding cavity became more sensitive to pushing and pulling of three more heavy fragments, the Rmsd values worsened. Even a stronger adjustment of the ring torso by positional constraints did not improve its orientation. This either could result form poor sampling or from unrealistic kinetics. Thus, we decided to dispense with a correct geometrical orientation of angles that involved hydrogen.

8.4.2 T4 lysozyme complexes

Dynamic self (re-)arrangement of a ring torso and one copy of each fragment type will be analyzed within the T4 lysozyme binding cavity (T4 lysozyme +ring torso +1 ethyl +1 methyl + 1 hydroxyl). Ethyl-benzene and phenol rings also (re)assembled to native like protein ligand complexes inside the hydrophobic T4 lysozyme cavity (see Fig. 8.4). The Rmsd of the fragment with respect to its native position shows phenol and ethyl-

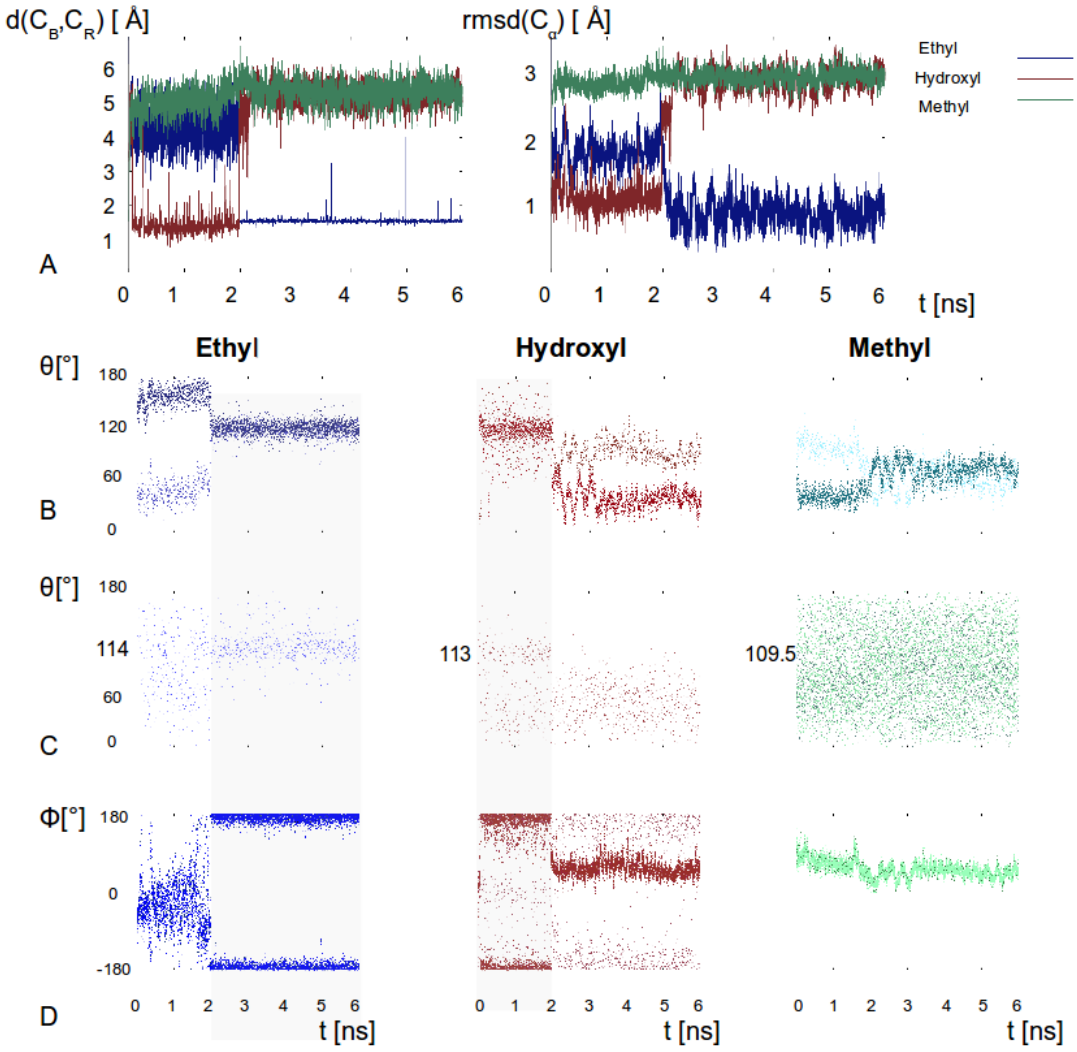


Figure 8.4: Ring torso in T4 lysozyme binding cavity with one methyl, ethyl and one hydroxyl group. While phenol is forming during the first 2 ns, (re-)arrangement results in an ethyl-benzene complex: (A) C_B,C_R distance versus simulation time (B)C_α-Rmsd versus simulation time with respect to native toluene, ethyl-benzene and phenol T4 lysozyme complex. Columns contain angle and dihedral angle data for ethyl, hydroxyl and methyl versus simulation time. Grey bars are indicating the correlation with native like fragment ring torso distance. (D) θ[°] 120° angles (E) θ[°] 114° 113° 109.5° angles: In case of hydroxyl and methyl these angles contain hydrogen. (E) Dihedral angles

benzene to adopt a native-like position inside the binding cavity. To account for the deformation of the cavity during simulations, the simulation data were aligned to the C_α atoms of the native structure within a 9$\AA$ shell around the ring. Correct distances do not necessarily correlate with low Rmsd values, indicating that non native binding modes are populated to a certain extent. Correct geometrical assembly of the ring torso and the fragments does not necessarily imply its correct orientation inside the binding pocket. However, rearrangement of the ring torso and the fragments is mostly resulting in the native binding mode. The geometrical orientation of angles (except the ones involving hydrogen see explanation above) is reproduced on average (see first

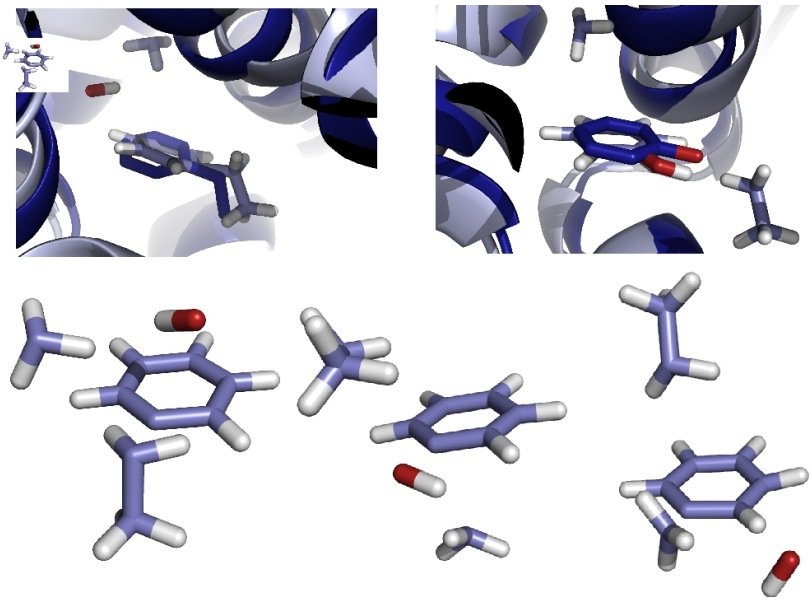


Figure 8.5: Superposition of native phenol and ethyl benzene in complex with T4 lysozyme with simulation arrangement

second and third panel in Fig. 8.4 angles and dihedral angles and snapshots in Fig. 8.5). A look at the dissociation constants in table 8.4 demonstrates the ethyl group to bind preferably to the ring torso, followed by the methyl and the hydroxyl group. Far away from being fully converged the simulations at least indicate this binding order.

	Benzene	Ethyl-benzene	Toulene	phenol
K_D(nM) kcal/mol	-5.8	-5.5	-5.3	-5.0

Table 8.4: Experimental binding free energies taken from [117]

8.5 Conclusion

This approach is limited by technical as well as physical problems: Very slow performance is based on the internal amber processing of external changes in the excluded list. From a quantum chemical perspective fragments as well as the ring torso are rough estimates of their models only. Manual preparation (antechamber and changes in topology) is slowing down performance. Inducing artificial charges on surfaces of ring and fragments means changing the electrostatic energy landscape of the whole system. This could especially affect dynamics in the T4 lysozyme the binding cavity. Additionally hydrogen involving angles poorly orientate correctly, resulting in a non chemical ring fragment assembly(see results and discussion). This protocol is a proof of principle and in its current form not suited for large fragment screening on a large scale. Future efforts should point into the direction of speeding up sampling. However it may serve as tool to identify possible alternative binding modes, distinguish to a certain extend the affinity of fragment complexes.

Chapter 9

Summary and Outlook

Proteins perform many complex tasks in living cells. Synthesized as amino acids chains, proteins adopt their well defined, complex, functional shape in a folding process. Since the function is coupled to their shape, the first part of this thesis aimed at exploring the conformational space by means of advanced sampling methods. Replica exchange simulations are a versatile tool to accelerate conformational sampling of biomolecules. Intrinsic properties of proteins entered into the construction of biasing potentials, to promote characteristic conformational changes, that regulate protein activity (for detailed description of the methodology refer to chapter 2). As shown in the second part of this thesis, enhanced sampling methods are also suited to study protein association. Detailed knowledge on protein peptide association gives molecular insight into binding modes of ligands and may even allow to discriminate between pharmaceutically relevant and irrelevant binders.

Summary

The biasing potential replica exchange method presented in chapter 3 was designed to specifically promote both, dihedral angle transitions along the polymer backbone and amino acid side chain flips. A two dimensional biasing potential penalizes Φ-Ψ dihedral angle combinations that populate distinct density peaks in the Ramachandran plot, corresponding to β-sheet or α-helical structures. Additionally, side chain flips around the χ_1 angle are enhanced by a one dimensional biasing potential to extend sampling of rotameric states. The replica framework is exploited optimally, by adjusting the level of the biasing potential on the fly in each replica. This resulted not only in high exchange rates between neighboring replicas and ensured occupancy/flow of all conformers in each replica but also guaranteed conformational variety. Canonical sampling remained unaffected after a short equilibration phase. Local dihedral angle transitions changed the global shape of three proteins/peptides that folded from random coils into their native structure more rapidly than in classical MD simulations of the same total length. As an advantage biasing potentials can be focused on selected protein regions only: Thus, an incorrect protein loop structure could fold into its native conformation rapidly.

Large scale domain-domain motions of proteins are essential for a wide range of biological processes. Principally, the biasing potential replica exchange method could be applied to flexible hinge regions that connect rigid protein domains. However, the performance improved by making use of elastic network models that predict protein dynamics on a large scale. The Hamiltonian replica exchange approach presented in chapter 4 combines this coarse grained information with the benefits of an atomistic simulation in explicit solvent, to enhance domain motions of proteins. Each protein is represented by a small number of centroids, corresponding to different protein domains. Biasing potentials act on the distances between these centroids and thus promote domain-domain motions of proteins. Potentials and distances are adjusted on the fly to account for both high exchange rates and good mixing of conformations among replicas. This ENM coupled approach enhanced global motions of two-domain segments of the glycoprotein 130 and of the protein cyanovirin-N significantly compared to classical MD simulations.

The biasing potential replica exchange method presented in chapter 3 is applied in chapter 5 to contribute towards understanding the mechanism of peptide loading to MHC class I molecules. As part of the viral immune response MHC class I molecules present antigenic peptides to cytotoxic T-cells at the cell surface. Peptide loading of MHC class I molecules in the endoplasmic reticulum can involve the chaperone protein tapasin. The human class I allotype HLA-B*44:02 with an aspartic acid at position 116 at the binding pocket floor loads peptides efficiently in the presence of tapasin only. In contrast, allotype HLA-B*44:05 differs in one single amino acid at position 116 (i.e tyrosine) but loads peptides independently of tapasin. Additionally, both allotypes have almost identical shape in complex with the same high affinity peptide. Hence, chapter 5 aims at elucidating the influence of amino acids at position 116 on the overall shape of both allotypes in the absence of bound peptide by means of free energy simulations: Increased conformational F pocket flexibility of B*44:02 allotypes was reflected in a low free energy barrier of F pocket opening. Moreover, it was associated with a side chain switch of the basic arginine 97 which may form salt bridges with nearby acidic aspartates at position 114, 156 and 116 close to the F pocket. Salt bridge rearrangement turned out to be coupled to the width of the B*44:02 F pocket. As arginine 97 did not interact with Tyr at position 116, the F pocket of the B*44:05 allotype barely changed its shape, reflected in a high energy barrier of F pocket opening. To establish a correlation between rotameric state of arginine 97 and the width of the F pocket, we made use of replica exchange simulations presented in chapter 3 to promote side chain flips of the switch and its interaction partners. While these induced flips provoked a widening of the F pocket of the B*44:02 allotype, they did not affect the F pocket shape of the B*44:05 allotype. Taken together, these results suggest an allosteric side chain switch that mediates the global dynamics of B*44:02 allotypes.

Identifying binding properties of small peptides or ligands to molecular targets is of interest not only to understand their function but also in the field of drug design. The second part of this thesis is devoted to the investigation of protein association by means of classical MD simulations and advanced sampling techniques.

In chapter 6 MD simulations and MMPB/SA calculations could give insight into the molecular mechanism of dipeptide binding to MHC class I molecules: High affinity peptides are ideally suited to speed up folding of MHC class I molecules in a step wise process: Thereby MHC class I molecules bind suboptimal low affinity peptides that are exchanged with higher affinity ones. As shown by experiments in the Springer lab, dipeptides such as GL fulfill the minimal requirements for MHC class I folding. GL is not only ideally suited to refold MHC class I molecules into their native form but also speeds up binding of high affinity peptides to MHC class I molecules. Classical MD simulations supported the hypothesis that GL binds to the C-terminal part of the binding pocket, acting as a placeholder for full length high affinity peptides. Dipeptides kept the binding pocket in a peptide receptive shape within simulation time. Moreover, MMPB/SA free energy calculations indicated higher binding affinity of GL compared to GG.

MMGB/SA (MMPB/SA) calculations estimate absolute binding free energies of protein ligand complexes, by post processing explicit solvent trajectories by means of generalized Born(GB) or Poisson Boltzmann implicit solvent models that account for solvent solute interactions. However, MMGB/SA calculations carry several sources of error that make absolute binding free energy an approximation of rigorous free energy values only. In chapter 7 rigorous relative free energy differences of different ligand molecules (alchemical transitions) are calculated based on the free energy perturbation method (FEP). The generalized Born implicit solvent model serves as basis for rigorous relative free energy calculations: Molecules are converted by scaling Lennard-Jones, Coulomb and Born radii parameters in a single-topology series of classical simulations. This series is integrated into a replica exchange scheme to speed up convergence. However, disappearing Born radii can increase the electrostatic reaction field (at constant partial charge of an atom) and provoke an increase of the electrostatic solvation of a disappearing atom. We circumvented these effects by scaling the charge quadratically combined with a linear change in Born radii to achieve a smooth free energy change along a coupling parameter of the alchemical transformation. The RE-FEP-GB approach was tested on the calculation of hydration free energies of amino acids, relative binding free energies of a ligand-receptor system and in silico alanine scanning of a peptide-protein complex. Despite of the inaccuracies that accompany the implicit GB model, calculated values are in good agreement with experimental data.

The RE-FEP-GB approach integrates solute solvent interactions by an implicit GB model, which ignores key water molecules and may fail predicting polar and buried charges as well. Thus, chapter 8 addresses an approach that mimics dynamic self (re-)arrangement of molecule fragments in MD simulations in explicit solvent. Surrounded by water molecules, these fragments can combine into realistic molecules that can degrade and (re-)arrange dynamically during a simulation. Thus, the chemical structure of ligands changes during MD simulations. In presence of a molecular target fragments are expected both, to form ligands and to adopt a realistic binding pose in the binding pocket. Lifetime of ligands in the desired binding pocket should ideally correspond to their binding affinities to reflect free energy differences (i.e. survival of high affinity ligands). Rearrangement of fragments is driven by bond, angle and dihedral angle potentials, that are effective if the fragments are in close proximity of

each other. Crystal structures of T4 lysozyme in complex with toluene, ethylbenzene and phenol as well as experimental data on the binding kinetics make this enzyme a suited molecular target, to prove the principle. The ligand was split into a ring and ethyl, methyl and hydroxyl fragments. While each fragment was interacting with the ring and the T4 lysozyme, different fragments did not interact with each other. Fragments and ring torso in explicit solution formed native-like toluene, ethyl-benzene and phenol rings. In solution with T4 lysozyme native-like phenol an ethyl-benzene complexes were sampled. Since this approach suffers from complicated manual fragment preparation and slow performance, future efforts are necessary to increase its efficiency.

Outlook

BP-REMD simulations successfully accelerated conformational sampling and folding of denatured proteins and proved suitable for loop refinement. Promoting local changes–namely the transitions of dihedral angle combinations along the polymer backbone and side chain flips–affected the global shape of proteins. This method could be extended on additional side chain dihedral angles, which could improve sampling of large side chains such as the one of tryptophan. Rarely observed in classical MD simulations, ω-backbone dihedral transitions of Proline that control cis trans transformations may become accessible. Allowing to penalize selected parts of the protein only, the BP-REMD approach could be optimized to refine unstructured protein loops in explicit solvent that are inaccessible for high resolution X-ray crystallography. Often incorrectly predicted by homology modeling procedures, this method could turn into a tool to accurately refine protein loops in explicit solvent.
Hamiltonian replica exchange combined with elastic network analysis enhanced global motions of protein domains. This approach could be extended to investigate quaternary structure (de-)formation of proteins consisting of not covalently bound domains of similar molecular weight. A combination with the BP-REMD method in chapter 4 in a 2 dimensional replica exchange scheme could be a promising extension to accelerate both local and global conformational changes at the same time.
In this thesis another development concerned the design of a protocol to dynamically adjust potential levels during a replica run. The replica exchange framework is optimally exploited by balancing high conformational variety and reasonable exchange rates. Refinement of this protocol should concern increments between neighboring replicas to additionally increase efficiency.
The RE-FEP-GB approach accurately characterizes protein complexes by rapidly calculating relative free energy changes associated with molecular alchemical transitions. Performance of this method depends on the accuracy of the generalized Born implicit solvent model and the efficiency of the annihilation creation pathway, that mimics alchemical transitions. Optimizing this pathway could result in a reduced number of parallel replicas and thus save computational power. A limited number of explicit water molecules–namely the ones, that are decisive for protein folding–could be integrated to both increase accuracy of calculations and to balance inaccurate implicit solvent effects. Moreover, graphical processing units (GPUs) are expected to drastically speed up RE-FEP-GB calculations to screen of ligands within minutes. Hence, RE-FEP-GB calculations could develop into an alternative, to rapidly perform rigorous free energy calculations.

The biasing potential Hamiltonian replica exchange (BP-REMD) presented in chapter 5 was restricted to disturb selected key side chains in the F pocket of MHC class I B*44:02 and B*44:05 alleles (neglecting the biasing of the backbone). Five replicas elucidated molecular amino acid side chain patterns to be coupled to global changes of F pocket shape and function. This method is expected to effectively and efficiently investigate the impact of local side chain flips on the overall shape of other types of proteins and systematically study global changes in proteins coupled to side chain motion.

Specific dynamics of peptides and proteins requires lots of problem solving strategies. In this thesis, advanced sampling techniques–mostly involving Hamiltonian replica exchange–could contribute to explore both conformational space of proteins and protein association in atomistic detail. These strategies were optimized to study conformational transitions, that regulate biological processes, as well as the shape and stability of protein complexes, allowing for a broad range of further applications.

Bibliography

[1] M. J. Abraham and J. E. Gready.
Ensuring mixing efficiency of replica-exchange molecular dynamics simulations.
J. Chem. Theory Comput., 4(7):1119–1128, 2008.

[2] E. J. Adams and A. M. Luoma.
The adaptable major histocompatibility complex (mhc) fold: structure and function of nonclassical and mhc class i-like molecules.
Annu. Rev. Immunol., 31:529–561, 2013.

[3] R. Affentranger, I. Tavernelli, and E. E. Di Iorio.
A novel hamiltonian replica exchange md protocol to enhance protein conformational space sampling.
J. Chem. Theory Comput., 2(2):217–228, 2006.

[4] C. B. Anfinsen.
Principles that govern the folding of protein chains.
Science, 181(4096):223–230, 1973.

[5] W. Astbury, S. Dickinson, and K. Bailey.
The x-ray interpretation of denaturation and the structure of the seed globulins.
Biochem. J., 29(10):2351–2360.1, October 1935.

[6] I. Bahar, A. R. Atilgan, and B. Erman.
Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential.
Folding and Design, 2(3):173–181, 1997.

[7] A. J. Ballard and C. Jarzynski.
Replica exchange with nonequilibrium switches.
Proc. Natl. Acad. Sci. U.S.A., 106(30):12224–12229, 2009.

[8] P. J. Barrett, J. Chen, M.-K. Cho, J.-H. Kim, Z. Lu, S. Mathew, D. Peng, Y. Song, W. D. Van Horn, T. Zhuang, F. D. SÃ¶nnichsen, and C. R. Sanders.
The quiet renaissance of protein nuclear magnetic resonance.
Biochemistry, 52(8):1303–1320, 2013.

[9] L. G. Barrientos, J. M. Louis, I. Botos, T. Mori, Z. Han, B. R. O'Keefe, M. R. Boyd, A. Wlodawer, and A. M. Gronenborn.
The domain-swapped dimer of cyanovirin-n is in a metastable folded state: Reconciliation of x-ray and {NMR} structures.
Structure, 10(5):673 – 686, 2002.

[10] M. A. Batalia and E. J. Collins.
Peptide binding by class i and class ii mhc molecules.
J. Pept. Sci., 43(4):281–302, 1997.

[11] A. Ben-Naim.
A farewell to entropy.
World Scientific, 2008.

[12] C. H. Bennett.
Efficient estimation of free energy differences from monte carlo data.
J. Comput. Phys., 22(2):245–268, 1976.

[13] C. A. Bewley, K. R. Gustafson, M. R. Boyd, D. G. Covell, A. Bax, G. M. Clore, and A. M. Gronenborn.
Solution structure of cyanovirin-n, a potent hiv-inactivating protein.
Nat. Struct. Mol. Biol., 5(7):571–578, 1998.

[14] T. Borchert, M. Mathieu, J. P. Zeelen, S. Courtneidge, and R. Wierenga.
The crystal structure of human csksh3: structural diversity near the rt-src and n-src loop.
FEBS lett., 341(1):79–85, 1994.

[15] I. Botos and A. Wlodawer.
Cyanovirin-n: a sugar-binding antiviral protein with a new twist.
Cell. Mol. Life Sci., 60(2):277–287, 2003.

[16] M. Bouvier.
Accessory proteins and the assembly of human class i mhc molecules: A molecular and structural perspective.
Mol. Immunol., 39:697 – 706, 2003.

[17] M. Bouvier and D. Wiley.
Structural characterization of a soluble and partial folded class i major histocompatibility chain β_2m heterodimer.
Nat. Struct. Mol. Biol., 5:377–384, 1998.

[18] J. Bravo, D. Staunton, J. K. Heath, and E. Jones.
Crystal structure of a cytokine-binding region of gp130.
EMBO J., 17(6):1665–1674, 1998.

[19] P. Brenner, C. R. Sweet, D. VonHandorf, and J. A. Izaguirre.
Accelerating the replica exchange method through an efficient all-pairs exchange.
J. Chem. Phys., 126(7):074103, 2007.

[20] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus.
Charmm: A program for macromolecular energy, minimization, and dynamics calculations.
J. Comput. Chem., 4(2):187–217, 1983.

[21] A. T. Brunger, P. D. Adams, and L. M. Rice.
Annealing in crystallography: a powerful optimization tool.
Prog. Biophys. Mol. Biol., 72(2):135 – 155, 1999.

[22] N.-V. Buchete and G. Hummer.
Peptide folding kinetics from replica exchange molecular dynamics.
Phys. Rev. E, 77:030902, Mar 2008.

[23] P. Burkhard, P. Taylor, and M. D. Walkinshaw.
X-ray structures of small ligand-fkbp complexes provide an estimate for hydrophobic interaction energies.
J. Mol. Biol., 295(4):953–962, 2000.

[24] F. Calvo.
All-exchanges parallel tempering.
J. Chem. Phys., 123(12):124106, 2005.

[25] R. A. Carr, M. Congreve, C. W. Murray, and D. C. Rees.
Fragment-based lead discovery: leads by design.
Drug Discov. Today, 10(14):987–992, 2005.

[26] D. Case, T. Darden, T. Cheatham III, C. Simmerling, J. Wang, R. Duke, R. Luo, R. Walker, W. Zhang, K. Merz, et al.
Amber 12.
San Francisco: University of California, 2012.

[27] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman.
Amber 12, 2012.

[28] D. Chandler.
Interfaces and the driving force of hydrophobic assembly.
Nature, 437(7059):640–647, 2005.

[29] Y. Chebaro, X. Dong, R. Laghaei, P. Derreumaux, and N. Mousseau.
Replica exchange molecular dynamics simulations of coarse-grained proteins in implicit solvent.
J. Phys. Chem. B, 113(1):267–274, 2009.

[30] Y. Chen and M. D. Barkley.
Toward understanding tryptophan fluorescence in proteins.
Biochemistry, 37(28):9976–9982, 1998.

[31] X. Cheng, G. Cui, V. Hornak, and C. Simmerling.
Modified replica exchange simulation methods for local structure refinement.
J. Phys. Chem. B, 109(16):8220–8230, 2005.

[32] C. D. Christ, A. E. Mark, and W. F. van Gunsteren.
Basic ingredients of free energy calculations: a review.
J. Comput. Chem., 31(8):1569–1582, 2010.

[33] P. Cossio, F. Marinelli, A. Laio, and F. Pietrucci.
Optimizing the performance of bias-exchange metadynamics: folding a 48-residue lysm domain using a coarse-grained model.

J. Phys. Chem. B, 114(9):3259–3265, 2010.

[34] J. Curuksu and M. Zacharias.
Enhanced conformational sampling of nucleic acids by a new hamiltonian replica exchange molecular dynamics approach.
J. Chem. Phys., 130(10):104110, 2009.

[35] T. Darden, D. York, and L. Pedersen.
Particle mesh ewald: An n$^a log(n)$ method for ewald sums in large systems.
J. Chem. Phys., 98:10089, 1993.

[36] X. Daura, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark.
Reversible peptide folding in solution by molecular dynamics simulation.
J. Mol. Biol., 280(5):925–932, 1998.

[37] A. de Ruiter and C. Oostenbrink.
Free energy calculations of protein–ligand interactions.
Curr. Biol. Chem., 15(4):547–552, 2011.

[38] Y. Deng and B. Roux.
Computations of standard binding free energies with molecular dynamics simulations.
J. Phys. Chem. B, 113(8):2234–2246, 2009.

[39] S. V. der Burg, M. Visseren, R. Brandt, W. M. Kast, and C. Melief.
Immunogenicity of peptides bound to mhc class i molecules depends on the mhc-peptide complex stability.
J. Immunol., 156(9):3308–3314, 1996.

[40] K. A. Dill.
Additivity principles in biochemistry.
J. Biological Chem., 272(2):701–704, 1997.

[41] J. Drenth.
Principles of protein X-ray crystallography.
Springer, 2007.

[42] R. O. Dror, R. M. Dirks, J. Grossman, H. Xu, and D. E. Shaw.
Biomolecular simulation: a computational microscope for molecular biology.
Annu. Rev. Biophys., 41:429–452, 2012.

[43] Y. Duan, C. Wu, S. Chowdhury, M. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, et al.
A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations.
J. Comput. Chem., 24(16):1999–2012, 2003.

[44] A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, et al.
Intrinsically disordered protein.
J. Mol. Graphics Modell., 19(1):26–59, 2001.

[45] A. K. Dunker, I. Silman, V. N. Uversky, and J. L. Sussman.
Function and structure of inherently disordered proteins.
Curr. Opin. Struct. Biol., 18(6):756–764, 2008.

[46] M. R. Eftink.
The use of fluorescence methods to monitor unfolding transitions in proteins.
Biophys. J., 66(2):482–501, 1994.

[47] T. Elliott.
How does tap associate with mhc class i molecules?
Immunol. today, 18(8):375–379, 1997.

[48] T. Elliott, V. Cerundolo, and A. Townsend.
Short peptides assist the folding of free class i heavy chains in solution.
Eur. J. Immunol., 22(12):3121–3125, 1992.

[49] T. J. Ewing, S. Makino, A. G. Skillman, and I. D. Kuntz.
Dock 4.0: search strategies for automated molecular docking of flexible molecule databases.
J. Comput.-Aided Mol. Des., 15(5):411–428, 2001.

[50] M. Fajer, D. Hamelberg, and J. A. McCammon.
Replica-exchange accelerated molecular dynamics (rexamd) applied to thermodynamic integration.
J. Chem. Theory Comput., 4(10):1565–1569, 2008.

[51] C. K. Fisher and C. M. Stultz.
Constructing ensembles for intrinsically disordered proteins.
Curr. Opin. Struct. Biol., 21(3):426–431, 2011.

[52] H. Fukunishi, O. Watanabe, and S. Takada.
On the hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction.
J. Chem. Phys., 116(20):9058–9067, 2002.

[53] E. Gallicchio, M. Lapelosa, and R. M. Levy.
Binding energy distribution analysis method (bedam) for estimation of protein-ligand binding affinities.
J. Chem. Theory Comput., 6(9):2961–2977, 2010.

[54] T. Garrett, M. Saper, P. Bjorkman, J. Strominger, and D. Wiley.
Specificity pockets for the side chains of peptide antigens in hla-aw68.
Nature, 342:692–696, 1989.

[55] M. A. Garstka, S. Fritzsche, I. Lenart, Z. Hein, G. Jankevicius, L. H. Boyle, T. Elliott, J. Trowsdale, A. N. Antoniou, M. Zacharias, and S. Springer.
Tapasin dependence of major histocompatibility complex class i molecules correlates with their conformational flexibility.
FASEB J., 25(11):3989–3998, 2011.

[56] H. Gohlke and D. A. Case.
Converging free energy estimates: Mm-pb (gb) sa studies on the protein-protein complex ras-raf.
J. Comput. Chem., 25(2):238–250, 2004.

[57] S. Gras, X. Saulquin, J.-B. Reiser, E. Debeaupuis, K. Echasserieau, A. Kissenpfennig, F. Legoux, A. Chouquet, M. Le Gorrec, P. Machillot, et al.
Structural bases for the affinity-driven selection of a public tcr against a dominant human cytomegalovirus epitope.

J. Immunol., 183(1):430–437, 2009.

[58] D. Gront and A. Kolinski.
Efficient scheme for optimization of parallel tempering monte carlo method.
J. Phys.: Condensed Matter, 19(3):036225, 2007.

[59] H. Grubmüller.
Predicting slow structural transitions in macromolecular systems: conformational flooding.
Phys. Rev. E, 52(3):2893, 1995.

[60] M. Gur, J. D. Madura, and I. Bahar.
Global transitions of proteins explored by a multiscale hybrid methodology: application to adenylate kinase.
Biophys. J., 105(7):1643–1652, 2013.

[61] D. Hamelberg, J. Mongan, and J. A. McCammon.
Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules.
J. Chem. Phys., 120(24):11919–11929, 2004.

[62] M. Harndahl, M. Rasmussen, G. Roder, I. Dalgaard Pedersen, M. Sørensen, M. Nielsen, and S. Buus.
Peptide-mhc class i stability is a better predictor than peptide affinity of ctl immunogenicity.
Eur. J. Immunol., 42(6):1405–1416, 2012.

[63] P. C. Heinrich, I. Behrmann, G. Müller-Newen, F. Schaper, and L. Graeve.
Interleukin-6-type cytokine signalling through the gp130 jak stat pathway.
Biochem. J., 334(2):297–314, 1998.

[64] B. Hess, H. Bekker, H. J. Berendsen, J. G. Fraaije, et al.
Lincs: a linear constraint solver for molecular simulations.
J. Comput. Chem., 18(12):1463–1472, 1997.

[65] K. Hinsen.
Analysis of domain motions by approximate normal mode calculations.
Proteins: Struct.; Funct.; and Genetics, 33(3):417–429, 1998.

[66] K. Hinsen, A.-J. Petrescu, S. Dellerue, M.-C. Bellissent-Funel, and G. R. Kneller.
Harmonicity in slow protein dynamics.
Chem. Phys., 261(1):25–37, 2000.

[67] R. W. Hockney.
Potential calculation and some applications.
Technical report, Langley Research Center, Hampton, Va., 1970.

[68] N. Homeyer and H. Gohlke.
Advances in molecular dynamics simulations and free energy calculations relevant for drug design.
Future Medicine, 2013.

[69] S. Honda, K. Yamasaki, Y. Sawada, and H. Morii.
10 residue folded peptide designed by segment statistics.
Structure, 12(8):1507–1518, Aug. 2004.

[70] B. Honig and A. Nicholls.
Classical electrostatics in biology and chemistry.
Science, 268:1144–1144, 1995.

[71] T. Hou, J. Wang, Y. Li, and W. Wang.
Assessing the performance of the mm/pbsa and mm/gbsa methods. 1. the accuracy of binding free energy calculations based on molecular dynamics simulations.
J. Chem. Inf. Model., 51(1):69–82, 2010.

[72] T. Hou, J. Wang, Y. Li, and W. Wang.
Assessing the performance of the molecular mechanics/poisson boltzmann surface area and molecular mechanics/generalized born surface area methods. ii. the accuracy of ranking poses generated from docking.
J. Comput. Chem., 32(5):866–877, 2011.

[73] M. Howarth, A. Williams, A. B. Tolstrup, and T. Elliott.
Tapasin enhances mhc class i peptide presentation according to peptide half-life.
Proc. Natl. Acad. Sci. U.S.A., 101(32):11737–11742, 2004.

[74] J. Hritz and C. Oostenbrink.
Hamiltonian replica exchange molecular dynamics using soft-core interactions.
J. Chem. Phys., 128(14):144121, 2008.

[75] X. Huang, M. Hagen, B. K. , R. A. Friesner, R. Zhou, and B. J. Berne.
Replica exchange with solute tempering: Efficiency in large scale systems.
J. Phys. Chem. B, 111(19):5405–5410, 2007.

[76] T. Huber, A. E. Torda, and W. F. van Gunsteren.
Structure optimization combining soft-core interaction functions, the diffusion equation method, and molecular dynamics.
J. Phys. Chem. A, 101(33):5926–5930, 1997.

[77] S. Hulpke, C. Baldauf, and R. Tampé.
Molecular architecture of the mhc i peptide-loading complex: one tapasin molecule is essential and sufficient for antigen processing.
FASEB J., 26(12):5071–5080, 2012.

[78] S. Hulpke and R. Tamp'e.
The mhc i loading complex: a multitasking machinery in adaptive immunity.
Trends Biochem. Sci., 38:412–420, 2013.

[79] S. G. Itoh and H. Okumura.
Coulomb replica-exchange method: Handling electrostatic attractive and repulsive forces for biomolecules.
J. Comput. Chem., 34(8):622–639, 2013.

[80] S. G. Itoh, H. Okumura, and Y. Okamoto.
Replica-exchange method in van der waals radius space: Overcoming steric restrictions for biomolecules.
J. Chem. Phys., 132(13):134105, 2010.

[81] A. K. Iversen, G. Stewart-Jones, G. H. Learn, N. Christie, C. Sylvester-Hviid, A. E. Armitage, R. Kaul, T. Beattie, J. K. Lee, Y. Li, et al.
Conflicting selective forces affect t cell receptor contacts in an immunodominant human immunodeficiency virus epitope.

Nat. Immunol., 7(2):179–189, 2006.

[82] W. Jiang, M. Hodoscek, and B. Roux.
Computation of absolute hydration and binding free energy with free energy perturbation distributed replica-exchange molecular dynamics.
J. Chem. Theory Comput., 5(10):2583–2588, 2009.

[83] W. Jiang and B. Roux.
Free energy perturbation hamiltonian replica-exchange molecular dynamics (fep/h-remd) for absolute ligand binding free energy calculations.
J. Chem. Theory Comput., 6(9):2559–2565, 2010.

[84] W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey, M. Klein, et al.
Comparison of simple potential functions for simulating liquid water.
J. Chem. Phys., 79(2):926, 1983.

[85] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives.
Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids.
J. Am. Chem. Soc., 118(45):11225–11236, 1996.

[86] J. Juraszek and P. Bolhuis.
Sampling the multiple folding mechanisms of trp-cage in explicit solvent.
Proc. Natl. Acad. Sci. U.S.A., 103(43):15859–15864, 2006.

[87] T. Kaihsu.
Conformational sampling for the impatient.
Biophys. Chem., 107(3):213–220, 2004.

[88] S. Kannan and M. Zacharias.
Enhanced sampling of peptide and protein conformations using replica exchange simulations with a peptide backbone biasing-potential.
Proteins: Struct., Funct., Bioinf., 66(3):697–706, 2007.

[89] S. Kannan and M. Zacharias.
Simulated annealing coupled replica exchange molecular dynamics-an efficient conformational sampling method.
J. Struct. Biol., 166(3):288–294, 2009.

[90] S. Kannan and M. Zacharias.
Application of biasing-potential replica-exchange simulations for loop modeling and refinement of proteins in explicit solvent.
Proteins: Struct., Funct., Bioinf., 78(13):2809–2819, 2010.

[91] M. Kara and M. Zacharias.
Influence of 8-oxoguanosine on the fine structure of dna studied with biasing-potential replica exchange simulations.
Biophys. J., 104(5):1089–1097, 2013.

[92] M. Karplus and J. A. McCammon.
Molecular dynamics simulations of biomolecules.
Nat. Struct. Mol. Biol., 9(9):646–652, Sept. 2002.

[93] J. Kästner.
Umbrella sampling.

Wiley Interdisciplinary Rev.: Comput. Mol. Sci., 1(6):932–942, 2011.

[94] I. V. Khavrutskii and A. Wallqvist.
Improved binding free energy predictions from single-reference thermodynamic integration augmented with hamiltonian replica exchange.
J. Chem. Theory Comput., 7(9):3001–3011, 2011.

[95] S. Kirkpatrick.
Optimization by simulated annealing: Quantitative studies.
J. Stat. Phys., 34(5-6):975–986, 1984.

[96] J. G. Kirkwood.
Statistical mechanics of fluid mixtures.
J. Chem. Phys., 3(5):300–313, 1935.

[97] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath.
Docking and scoring in virtual screening for drug discovery: methods and applications.
Nat. Rev. Drug Discovery, 3(11):935–949, 2004.

[98] P. Kollman.
Free energy calculations: applications to chemical and biochemical phenomena.
Chem. Rev.s, 93(7):2395–2417, 1993.

[99] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, et al.
Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models.
Acc. Chem. Res., 33(12):889–897, 2000.

[100] J. Kostrowicki and H. A. Scheraga.
Application of the diffusion equation method for global optimization to oligopeptides.
J. Phys. Chem., 96(18):7442–7449, 1992.

[101] V. Kräutler, W. F. van Gunsteren, and P. H. Hünenberger.
A fast shake algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations.
J. Comput. Chem., 22(5):501–508, 2001.

[102] M. B. Kubitzki and B. L. de Groot.
The atomistic mechanism of conformational transition in adenylate kinase: A tee-rex molecular dynamics study.
Structure, 16(8):1175–1182, 2008.

[103] B. Kuhn, P. Gerber, T. Schulz-Gasch, and M. Stahl.
Validation and use of the mm-pbsa approach for drug discovery.
J. Med. Chem., 48(12):4040–4048, 2005.

[104] S. Kumar, J. Rosenberg, D. Bouzida, R. Swendsen, and P. Kollman.
The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method.
J. Comput. Chem., 13(8):1011–1021, 1992.

[105] E. Kurimoto, K. Kuroki, Y. Yamaguchi, M. Yagi-Utsumi, T. Igaki, T. Iguchi, K. Maenaka, and K. Kato.
Structural and functional mosaic nature of mhc class i molecules in their peptide-free form.
Mol. Immunol., 55(3):393–399, 2013.

[106] A. Laio and M. Parrinello.
Escaping free-energy minima.
Proc. Natl. Acad. Sci. U.S.A., 99(20):12562–12566, 2002.

[107] O. F. Lange, L. V. Schäfer, and H. Grubmüller.
Flooding in gromacs: Accelerated barrier crossings in molecular dynamics.
J. Comput. Chem., 27(14):1693–1702, 2006.

[108] M. Lapelosa, E. Gallicchio, and R. M. Levy.
Conformational transitions and convergence of absolute binding free energy calculations.
J. Chem. Theory Comput., 8(1):47–60, 2012.

[109] A. J. Lee and S. W. Rick.
Improving replica exchange using driven scaling.
J. Chem. Phys., 131(17):174113, 2009.

[110] P. E. Leopold, M. Montal, and J. N. Onuchic.
Protein folding funnels: a kinetic approach to the sequence-structure relationship.
Proc. Natl. Acad. Sci. U.S.A., 89(18):8721–8725, 1992.

[111] J. W. Lewis, A. Neisig, J. Neefjes, and T. Elliott.
Point mutations in the α_2 domain of hla-a2.1 define a functionally relevant interaction with tap.
Curr. Biol., 6(7):873–883, 1996.

[112] C. Li, M. Pazgier, C. Li, W. Yuan, M. Liu, G. Wei, W.-Y. Lu, and W. Lu.
Systematic mutational analysis of peptide inhibition of the p53–mdm2/mdmx interactions.
J. Mol. Biol., 398(2):200–213, 2010.

[113] C. Li, M. Pazgier, M. Liu, W.-Y. Lu, and W. Lu.
Apamin as a template for structure-based rational design of potent peptide activators of p53.
Angew. Chem., 121(46):8868–8871, 2009.

[114] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw.
How fast-folding proteins fold.
Science, 334(6055):517–520, 2011.

[115] K. Lindorff-Larsen, N. Trbovic, P. Maragakis, S. Piana, and D. E. Shaw.
Structure and dynamics of an unfolded protein examined by molecular dynamics simulation.
J. Am. Chem. Soc., 134(8):3787–3791, 2012.

[116] M. Lingenheil, R. Denschlag, G. Mathias, and P. Tavan.
Efficiency of exchange schemes in replica exchange.
Chem. Phys. Lett., 478(1â3):80–84, 2009.

[117] L. Liu, W. A. Baase, M. M. Michael, and B. W. Matthews.
Use of stabilizing mutations to engineer a charged group within a ligand-binding hydrophobic cavity in t4 lysozyme.
Biochemistry, 48(37):8842–8851, 2009.

[118] P. Liu, X. Huang, R. Zhou, and B. J. Berne.
Hydrophobic aided replica exchange: an efficient algorithm for protein folding in explicit solvent.
J. Phys. Chem. B, 110(38):19018–19022, 2006.

[119] P. Liu, B. Kim, R. A. Friesner, and B. J. Berne.
Replica exchange with solute tempering: A method for sampling biological systems in explicit water.
Proc. Natl. Acad. Sci. U.S.A. of the United States of America, 102(39):13749–13754, 2005.

[120] M. P. Luitz and M. Zacharias.
Role of tyrosine hot-spot residues at the interface of colicin e9 and immunity protein 9: A comparative free energy simulation study.
Proteins: Struct., Funct., Bioinf., 81(3):461–468, 2013.

[121] E. Lyman and D. M. Zuckerman.
Resolution exchange simulation with incremental coarsening.
J. Chem. Theory Comput., 2(3):656–666, 2006.

[122] W. Macdonald, A. Purcell, N. Mifsud, L. Ely, D. Williams, L. Chang, J. Gorman, C. Clements, L. Kjer-Nielsen, D. Koelle, et al.
A naturally selected dimorphism within the hla-b44 supertype alters class i structure, peptide repertoire, and t cell recognition.
J. Exp. Med., 198(5):679, 2003.

[123] J. Machta.
Strengths and weaknesses of parallel tempering.
Phys. Rev. E, 80:056706, Nov 2009.

[124] D. R. Madden.
The three-dimensional structure of peptide-mhc complexes.
Annu. Rev. Immunol., 13(1):587–622, 1995.

[125] L. V. Mancour, H. N. Daghestani, S. Dutta, G. H. Westfield, J. Schilling, A. N. Oleskie, J. F. Herbstman, S. Z. Chou, and G. Skiniotis.
Ligand-induced architecture of the leptin receptor signaling complex.
Mol. Cell, 48(4):655–661, 2012.

[126] R. Matadeen, W.-C. Hon, J. K. Heath, E. Y. Jones, and S. Fuller.
The dynamics of signal triggering in a gp130-receptor complex.
Structure, 15(4):441–448, 2007.

[127] A. May and M. Zacharias.
Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein–protein docking.
Proteins: Struct., Funct., Bioinf., 70(3):794–80–, 2008.

[128] J. A. McCammon.

Dynamics of folded proteins.
Nature, 267:16, 1977.

[129] Y. Meng, D. Sabri Dashti, and A. E. Roitberg.
Computing alchemical free energy differences with hamiltonian replica exchange molecular dynamics (h-remd) simulations.
J. Chem. Theory Comput., 7(9):2721–2727, 2011.

[130] J. Michel and J. W. Essex.
Hit identification and binding mode predictions by rigorous free energy simulations.
J. Med. Chem., 51(21):6654–6664, 2008.

[131] J. Michel, M. L. Verdonk, and J. W. Essex.
Protein-ligand binding affinity predictions by implicit solvent simulations:âa tool for lead optimization?
J. Med. Chem., 49(25):7427–7439, 2006.

[132] B. R. Miller, T. D. McGee Jr, J. M. Swails, N. Homeyer, H. Gohlke, and A. E. Roitberg.
Mmpbsa. py: An efficient program for end-state free energy calculations.
J. Chem. Theory Comput., 8(9):3314–3321, 2012.

[133] S. Miyamoto and P. A. Kollman.
Settle: an analytical version of the shake and rattle algorithm for rigid water models.
J. Comput. Chem., 13(8):952–962, 1992.

[134] S. L. C. Moors, S. Michielssens, and A. Ceulemans.
Improved replica exchange method for native-state protein sampling.
J. Chem. Theory Comput., 7(1):231–237, 2011.

[135] Y. Mu.
Dissociation aided and side chain sampling enhanced hamiltonian replica exchange.
J. Chem. Phys., 130(16):164107–164107, 2009.

[136] K. P. Murphy, D. Xie, K. S. Thompson, L. M. Amzel, and E. Freire.
Entropy in biological binding processes: estimation of translational entropy loss.
Proteins: Struct., Funct., Bioinf., 18(1):63–67, 1994.

[137] W. Nadler and U. H. E. Hansmann.
Optimized explicit-solvent replica exchange molecular dynamics from scratch.
J. Phys. Chem. B, 112(34):10386–10387, 2008.
PMID: 18671362.

[138] W. Nadler, J. H. Meinke, and U. H. E. Hansmann.
Folding proteins by first-passage-times-optimized replica exchange.
Phys. Rev. E, 78:061905, Dec 2008.

[139] J. Neefjes, M. L. Jongsma, P. Paul, and O. Bakke.
Towards a systems understanding of mhc class i and mhc class ii antigen presentation.
Nat. Rev. Immunol., 11(12):823–836, 2011.

[140] J. W. Neidigh, R. M. Fesinmeyer, and N. H. Andersen.
Designing a 20-residue protein.
Nat. Struct. Mol. Biol., 9(6):425–430, June 2002.

[141] A. Nicholls and B. Honig.
A rapid finite difference algorithm, utilizing successive over-relaxation to solve the poisson–boltzmann equation.
J. Comput. Chem., 12(4):435–445, 1991.

[142] M. Nina, W. Im, and B. Roux.
Optimized atomic radii for protein continuum electrostatics solvation forces.
Biophys. Chem., 78(1):89–96, 1999.

[143] H. Nymeyer and A. E. García.
Simulation of the folding equilibrium of α-helical peptides: A comparison of the generalized born approximation with explicit solvent.
Proc. Natl. Acad. Sci. U.S.A., 100(24):13934–13939, 2003.

[144] T. Okabe, M. Kawata, Y. Okamoto, and M. Mikami.
Replica-exchange monte carlo method for the isobaric-isothermal ensemble.
Chem. Phys. Lett., 335(5):435–439, 2001.

[145] Y. Okamoto.
Generalized-ensemble algorithms: enhanced sampling techniques for monte carlo and molecular dynamics simulations.
J. Mol. Graphics Modell., 22(5):425–439, 2004.

[146] A. Okur, D. R. Roe, G. Cui, V. Hornak, and C. Simmerling.
Improving convergence of replica-exchange simulations through coupling to a high-temperature structure reservoir.
J. Chem. Theory Comput., 3(2):557–568, 2007.

[147] M. A. Olson, M. Feig, and C. L. Brooks.
Prediction of protein loop conformations using multiscale modeling methods with physical energy scoring functions.
J. Comput. Chem., 29(5):820–831, 2008.

[148] R. Olson, K. E. Huey-Tubman, C. Dulac, and P. J. Bjorkman.
Structure of a pheromone receptor-associated mhc molecule with an open and empty groove.
PLoS Biol., 3(8):e257, 2005.

[149] A. Onufriev, D. Bashford, and D. A. Case.
Modification of the generalized born model suitable for macromolecules.
J. Phys. Chem. B, 104(15):3712–3720, 2000.

[150] A. Onufriev, D. Bashford, and D. A. Case.
Exploring protein native states and large-scale conformational changes with a modified generalized born model.
Proteins: Struct., Funct., Bioinf., 55(2):383–394, 2004.

[151] B. Ortmann, M. J. Androlewicz, and P. Cresswell.
Mhc class i/β2-microglobulin complexes associate with tap transporters before peptide binding.
Nature, 368:864–867, 1994.

[152] K. Ostermeir and M. Zacharias.
Advanced replica-exchange sampling to study the flexibility and plasticity of peptides and proteins.
Biochim. Biophys. Acta (BBA)-Proteins and Proteomics, 1834(5):847–853, 2013.

[153] K. Ostermeir and M. Zacharias.
Hamiltonian replica-exchange simulations with adaptive biasing of peptide backbone and side chain dihedral angles.
J. Comput. Chem., 35(2):150–158, 2014.

[154] B. Park, S. Lee, E. Kim, and K. Ahn.
A single polymorphic residue within the peptide-binding cleft of mhc class i molecules determines spectrum of tapasin dependence.
J. Immunol., 170(2):961–968, 2003.

[155] H. Park, J. Lee, and S. Lee.
Critical assessment of the automated autodock as a new docking tool for virtual screening.
Proteins: Struct., Funct., Bioinf., 65(3):549–554, 2006.

[156] L. Pauling, R. B. Corey, and H. R. Branson.
The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain.
Proc. Natl. Acad. Sci. U.S.A. of the United States of America, 37(4):205, 1951.

[157] M. Pazgier, M. Liu, G. Zou, W. Yuan, C. Li, C. Li, J. Li, J. Monbo, D. Zella, and S. G. Tarasov.
Structural basis for high-affinity peptide inhibition of p53 interactions with mdm2 and mdmx.
Proc. Natl. Acad. Sci. U.S.A., 106(12):4665–4670, 2009.

[158] X. Periole and A. E. Mark.
Convergence and sampling efficiency in replica exchange simulations of peptide folding in explicit solvent.
J. Chem. Phys., 126(1):014903, 2007.

[159] S. Piana and A. Laio.
A bias-exchange approach to protein folding.
J. Phys. Chem. B, 111(17):4553–4559, 2007.

[160] S. Piana, K. Lindorff-Larsen, and D. E. Shaw.
Protein folding kinetics and thermodynamics from atomistic simulation.
Proc. Natl. Acad. Sci. U.S.A., 109(44):17845–17850, 2012.

[161] J. W. Ponder, D. A. Case, et al.
Force fields for protein simulations.
Adv. Prot. Chem., 66:27–86, 2003.

[162] W. Pos, D. K. Sethi, M. J. Call, M.-S. E. Schulze, A.-K. Anders, J. Pyrdol, and K. W. Wucherpfennig.
Crystal structure of the hla-dm–hla-dr1 complex defines mechanisms for rapid peptide selection.
Cell, 151(7):1557–1568, 2012.

[163] P. Praveen, R. Yaneva, H. Kalbacher, and S. Springer.
Tapasin edits peptides on mhc class i molecules by accelerating peptide exchange.
Eur. J. Immunol., 40(1):214–224, 2010.

[164] C. Predescu, M. Predescu, and C. V. Ciobanu.
On the efficiency of exchange in parallel tempering monte carlo simulations.
J. Phys. Chem. B, 109(9):4189–4196, 2005.

[165] L. Qiu, S. A. Pabit, A. E. Roitberg, and S. J. Hagen.
Smaller and faster: the 20-residue trp-cage protein folds in 4 μs.
J. Am. Chem. Soc., 124(44):12952–12953, 2002.

[166] G. Ramachandran, C. t. Ramakrishnan, and V. Sasisekharan.
Stereochemistry of polypeptide chain configurations.
J. Mol. Biol., 7(1):95–99, 1963.

[167] X. Rao, A. I. C. F. Costa, D. van Baarle, and C. Keşmir.
A comparative study of hla binding affinity and ligand diversity: implications for generating immunodominant cd8+ t cell responses.
J. Immunol., 182(3):1526–1532, 2009.

[168] D. C. Rees, M. Congreve, C. W. Murray, and R. Carr.
Fragment-based lead discovery.
Nat. Rev. Drug Discovery, 3(8):660–672, 2004.

[169] R. Riemann and M. Zacharias.
Reversible scaling of dihedral angle barriers during molecular dynamics to improve structure prediction of cyclic peptides.
J. Peptide Res., 63(4):354–364, 2004.

[170] R. N. Riemann and M. Zacharias.
Refinement of protein cores and protein peptide interfaces using a potential scaling approach.
Protein Eng. Des. Sel., 18(10):465–476, October 2005.

[171] P. T. Robert Denschlag, Martin Lingenheil.
Efficiency reduction and pseudo-convergence in replica exchange sampling of peptide folding-unfolding equilibria.
Chem. Phys. Lett., 458(1â3):244–248, 2008.

[172] T. Rodinger, P. L. Howell, and R. Pomès.
Calculation of absolute protein-ligand binding free energy using distributed replica sampling.
J. Chem. Phys., 129(15):155102, 2008.

[173] A. E. Roitberg, A. Okur, and C. Simmerling.
Coupling of replica exchange simulations to a non-boltzmann structure reservoir.
J. Phys. Chem. B, 111(10):2415–2418, 2007.

[174] E. Rosta, N.-V. Buchete, and G. Hummer.
Thermostat artifacts in replica exchange molecular dynamics simulations.
J. Chem. Theory Comput., 5(5):1393–1399, 2009.

[175] E. Rosta and G. Hummer.
Error and efficiency of simulated tempering simulations.

J. Chem. Phys., 132(3):034102, 2010.

[176] B. Roux.
The calculation of the potential of mean force using computer simulations.
Comput. Phys. Commun., 91(1â3):275–282, 1995.

[177] R. Roy, S. Hohng, and T. Ha.
A practical guide to single-molecule fret.
Nat. Methods, 5(6):507–516, 2008.

[178] J. Z. Ruscio, N. L. Fawzi, and T. Head-Gordon.
How hot? systematic convergence of the replica exchange method using multiple reservoirs.
J. Comput. Chem., 31(3):620–627, 2010.

[179] B. Sadasivan, P. J. Lehner, B. Ortmann, T. Spies, and P. Cresswell.
Roles for calreticulin and a novel glycoprotein, tapasin, in the interaction of mhc class i molecules with tap.
Immunity, 5(2):103–114, 1996.

[180] S. K. Saini, E. T. Abualrous, A.-S. T. Tigan, K. Covella, U. Wellbrock, and S. Springer.
Not all empty mhc class i molecules are molten globules: Tryptophan fluorescence reveals a two-step mechanism of thermal denaturation.
Mol. Immunol., 54(3):386–396, 2013.

[181] S. K. Saini, K. Ostermeir, V. R. Ramnarayan, H. Schuster, M. Zacharias, and S. Springer.
Dipeptides promote folding and peptide binding of mhc class i molecules.
Proc. Natl. Acad. Sci. U.S.A., 110(38):15383–15388, 2013.

[182] C. Schneeweiss, M. Garstka, J. Smith, M.-T. Hütt, and S. Springer.
The mechanism of action of tapasin in the peptide exchange on mhc class i molecules determined from kinetics simulation studies.
Mol. Immunol., 46(10):2054 – 2063, 2009.

[183] A. Schütt, M. Zacharias, N. Schneider, S. Horn, J. Götzinger, S. Rose John, and D. Schmidt Arras.
gp130 activation is regulated by d2-d3 interdomain connectivity.
Biochem. J., 450(3):487–496, 2013.

[184] M. M. Seibert, A. Patriksson, B. Hess, and D. van der Spoel.
Reproducible polypeptide folding and structure prediction using molecular dynamics simulations.
J. Mol. Biol., 354(1):173–183, 2005.

[185] D. K. Shenfeld, H. Xu, M. P. Eastwood, R. O. Dror, and D. E. Shaw.
Minimizing thermodynamic length to select intermediate states for free-energy calculations and replica-exchange simulations.
Phys. Rev. E, 80:046705, Oct 2009.

[186] Z. Shi, C. A. Olson, G. D. Rose, R. L. Baldwin, and N. R. Kallenbach.
Polyproline ii structure in a sequence of seven alanine residues.
Proc. Natl. Acad. Sci. U.S.A., 99(14):9190–9195, 2002.

[187] M. R. Shirts and V. S. Pande.
Solvation free energies of amino acid side chain analogs for common molecular mechanics water models.
J. Chem. Phys., 122(13):134508, 2005.

[188] D. Shivakumar, Y. Deng, and B. Roux.
Computations of absolute solvation free energies of small molecules using explicit and implicit solvent model.
J. Chem. Theory Comput., 5(4):919–930, 2009.

[189] F. Sieker, S. Springer, and M. Zacharias.
Comparative molecular dynamics analysis of tapasin-dependent and -independent mhc class i alleles.
Protein Sci., 16(2):299–308, 2007.

[190] M. L. Silver, K. C. Parker, and D. C. Wiley.
Reconstitution by mhc-restricted peptides of hla-a2 heavy chain with β2-microglobulin, in vitro.
Nature, 350:619–622, 1991.

[191] C. Simmerling, J. L. Miller, and P. A. Kollman.
Combined locally enhanced sampling and particle mesh ewald as a strategy to locate the experimental structure of a nonhelical nucleic acid.
J. Am. Chem. Soc., 120(29):7149–7155, 1998.

[192] D. Sindhikara, Y. Meng, and A. E. Roitberg.
Exchange frequency in replica exchange molecular dynamics.
J. Chem. Phys., 128(2):024103, 2008.

[193] D. J. Sindhikara, D. J. Emerson, and A. E. Roitberg.
Exchange often and properly in replica exchange molecular dynamics.
J. Chem. Theory Comput., 6(9):2804–2808, 2010.

[194] D. Sitkoff, K. A. Sharp, and B. Honig.
Accurate calculation of hydration free energies using macroscopic solvent models.
J. Phys. Chem., 98(7):1978–1988, 1994.

[195] J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman, and D. A. Case.
Continuum solvent studies of the stability of dna, rna, and phosphoramidate-dna helices.
J. Am. Chem. Soc., 120(37):9401–9409, 1998.

[196] T. Steinbrecher and A. Labahn.
Towards accurate free energy calculations in ligand protein-binding studies.
Curr. Med. Chem., 17(8):767–785, 2010.

[197] D. Steiner, C. Oostenbrink, F. Diederich, M. Zürcher, and W. F. van Gunsteren.
Calculation of binding free energies of inhibitors to plasmepsin ii.
J. Comput. Chem., 32(9):1801–1812, 2011.

[198] L. J. Stern and D. C. Wiley.
Antigenic peptide binding by class i and class {II} histocompatibility proteins.
Structure, 2(4):245–251, 1994.

[199] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson.
Semianalytical treatment of solvation for molecular mechanics and dynamics.
J. Am. Chem. Soc., 112(16):6127–6129, 1990.

[200] Y. Sugita, A. Kitao, and Y. Okamoto.
Multidimensional replica-exchange method for free-energy calculations.
J. Chem. Phys., 113(15):6042–6051, 2000.

[201] Y. Sugita and Y. Okamoto.
Replica-exchange molecular dynamics method for protein folding.
Chem. Phys. Lett., 314(1â2):141–151, 1999.

[202] W.-K. Suh, M. A. Derby, M. F. Cohen-Doyle, G. J. Schoenhals, K. Früh, J. A. Berzofsky, and D. B. Williams.
Interaction of murine mhc class i molecules with tapasin and tap enhances peptide loading and involves the heavy chain α3 domain.
J. Immunol., 162(3):1530–1540, 1999.

[203] R. H. Swendsen and J.-S. Wang.
Replica monte carlo simulation of spin glasses.
Phys. Rev. Lett., 57(21):2607–2609, 1986.

[204] F. Tama and Y.-H. Sanejouand.
Conformational change of proteins arising from normal mode calculations.
Protein Eng., 14(1):1–6, 2001.

[205] C. Tan, Y.-H. Tan, and R. Luo.
Implicit nonpolar solvent models.
J. Phys. Chem. B, 111(42):12263–12274, 2007.

[206] K. Tappura, M. Lahtela-Kakkonen, and O. Teleman.
A new soft-core potential function for molecular dynamics applied to the prediction of protein loop conformations.
J. Comput. Chem., 21(5):388–397, 2000.

[207] M. M. Tirion.
Large amplitude elastic motions in proteins from a single-parameter, atomic analysis.
Phys. Rev. Lett., 77(9):1905, 1996.

[208] D. Tobi and I. Bahar.
Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state.
Proc. Natl. Acad. Sci. U.S.A. of the United States of America, 102(52):18908–18913, 2005.

[209] G. M. Torrie and J. P. Valleau.
Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling.
J. Comput. Phys., 23(2):187–199, 1977.

[210] A. Townsend, T. Elliott, V. Cerundolo, L. Foster, B. Barber, and A. Tse.
Assembly of mhc class i molecules analyzed in vitro.
Cell, 62(2):285–295, 1990.

[211] S. Trebst, M. Troyer, and U. H. E. Hansmann.
Optimized parallel tempering simulations of proteins.
J. Chem. Phys., 124(17):174903, 2006.

[212] H. Turnquist, S. Vargas, E. Schenk, M. McIlhaney, A. Reber, and J. Solheim.
The interface between tapasin and mhc class i.
Immunol. Res., 25:261–269, 2002.

[213] W. F. van Gunsteren, X. Daura, and A. E. Mark.
Gromos force field.
Encyclopedia Comput. Chem., 1998.

[214] W. F. van Gunsteren, J. Dolenc, and A. E. Mark.
Molecular simulation as an aid to experimentalists.
Curr. Opin. Struct. Biol., 18(2):149–153, 2008.

[215] P. J. Van Laarhoven and E. H. Aarts.
Simulated annealing.
Springer, 1987.

[216] L. Verlet.
Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules.
Phys. Rev., 159(1):98, 1967.

[217] J. Vreede, M. G. Wolf, S. W. de Leeuw, and P. G. Bolhuis.
Reordering hydrogen bonds using hamiltonian replica exchange enhances sampling of conformational changes in biomolecular systems.
J. Phys. Chem. B, 113(18):6484–6494, 2009.

[218] J. Wang, W. Zhu, G. Li, and U. H. E. Hansmann.
Velocity-scaling optimized replica exchange molecular dynamics of proteins in a hybrid explicit/implicit solvent.
J. Chem. Phys., 135(8):084115, 2011.

[219] P. Wearsch, Pamela A et Cresswell.
Selective loading of high-affinity peptides onto major histocompatibility complex class i molecules by the tapasin-erp57 heterodimer.
Nat. Immunol., 8:1529–2908, 2007.

[220] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner.
A new force field for molecular mechanical simulation of nucleic acids and proteins.
J. Am. Chem. Soc., 106(3):765–784, 1984.

[221] A. P. Williams, C. A. Peh, A. W. Purcell, J. McCluskey, and T. Elliott.
Optimization of the mhc class i peptide cargo is dependent on tapasin.
Immunity, 16(4):509–520, 2002.

[222] R. Wolfenden, L. Andersson, P. Cullis, and C. Southgate.
Affinities of amino acid side chains for solvent water.
Biochemistry, 20(4):849–855, 1981.

[223] C. J. Woods, J. W. Essex, and M. A. King.
The development of replica-exchange-based free-energy methods.
J. Phys. Chem. B, 107(49):13703–13710, 2003.

[224] C. Wright, P. Kozik, M. Zacharias, and S. Springer.
Tapasin and other chaperones: models of the mhc class i loading complex.
Biol. Chem., 385:763–861, 2004.

[225] X. Wu, M. Hodoscek, and B. R. Brooks.
Replica exchanging self-guided langevin dynamics for efficient and accurate conformational sampling.
J. Chem. Phys., 137(4):044106, 2012.

[226] K. Wüthrich.
Protein structure determination in solution by nuclear magnetic resonance spectroscopy.
Science, 243(4887):45–50, 1989.

[227] W. Xu, T. Lai, Y. Yang, and Y. Mu.
Reversible folding simulation by hybrid hamiltonian replica exchange.
J. Chem. Phys., 128(17):175105, 2008.

[228] F. Yang, C. A. Bewley, J. M. Louis, K. R. Gustafson, M. R. Boyd, A. M. Gronenborn, G. M. Clore, and A. Wlodawer.
Crystal structure of cyanovirin-n, a potent hiv-inactivating protein, shows unexpected domain swapping.
J. Mol. Biol., 288(3):403–412, 1999.

[229] M. Zacharias.
Combining elastic network analysis and molecular dynamics simulations by hamiltonian replica exchange.
J. Chem. Theory Comput., 4(3):477–487, 2008.

[230] M. Zacharias and S. Springer.
Conformational flexibility of the mhc class i α_1-α_2 domain in peptide bound and free states: A molecular dynamics simulation study.
Biophys. J., 87(4):2203–2214, 2004.

[231] D. Zernich, A. Purcell, W. Macdonald, L. Kjer-Nielsen, L. Ely, N. Laham, T. Crockford, N. Mifsud, M. Bharadwaj, L. Chang, et al.
Natural hla class i polymorphism controls the pathway of antigen presentation and susceptibility to viral evasion.
J. Exp. Med., 200(1):13–24, 2004.

[232] C. Zhang, A. Anderson, and C. DeLisi.
Structural principles that govern the peptide-binding motifs of class i mhc molecules.
J. Mol. Biol., 281(5):929–947, 1998.

[233] Z. Zhang, Y. Shi, and H. Liu.
Molecular dynamics simulations of peptides and proteins with amplified collective motions.
Biophys. J., 84(6):3583–3593, 2003.

[234] W. Zheng, M. Andrec, E. Gallicchio, and R. M. Levy.
Simulating replica exchange simulations of protein folding with a kinetic network model.
Proc. Natl. Acad. Sci. U.S.A., 104(39):15340–15345, 2007.

[235] W. Zheng, M. Andrec, E. G. R. M, and Levy.
Simple continuous and discrete models for simulating replica exchange simulations of protein folding.
J. Phys. Chem. B, 112(19):6083–6093, 2008.

[236] D. M. Zuckerman and E. Lyman.
A second look at canonical sampling of biomolecules using replica exchange simulation.
J. Chem. Theory Comput., 2(4):1200–202, 2006.

[237] R. Zwanzig, A. Szabo, and B. Bagchi.
Levinthal's paradox.
Proc. Natl. Acad. Sci. U.S.A., 89(1):20–22, 1992.

[238] R. W. Zwanzig.
High-temperature equation of state by a perturbation method i nonpolar gases.
J. Chem. Phys., 22(8):1420–1426, 1954.

List of publications

Parts of this thesis have been published in or have been submitted to peer-reviewed journals. Manuscripts for contributions in peer-reviewed journals are in preparation:

- Katja Ostermeir, Martin Zacharias
  Advanced replica-exchange sampling to study the flexibility and plasticity of peptides and proteins
  Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 1834(5):847-853, 2013.

- Sunil Kumar Saini, Katja Ostermeir, Venkat Raman Ramnarayan, Heiko Schuster, Martin Zacharias, Sebastian Springer.
  Dipeptides promote folding and peptide binding of mhc class i molecules.
  Proceedings of the National Academy of Sciences, 110(38):15383-15388, 2013.

- Katja Ostermeir Martin Zacharias.
  Hamiltonian replica-exchange simulations with adaptive biasing of peptide backbone and side chain dihedral angles
  Journal of computational chemistry, 35(2):150-158, 2014

- Katja Ostermeir, Sebastian Springer, Martin Zacharias.
  *Allosteric coupling between side chain interactions and binding pocket flexibility in HLA-B*44:02 molecules investigated by molecular dynamics simulations.*
  accepted in Molecular Immunology

- Katja Ostermeir, Martin Zacharias.
  Hamiltonian replica-exchange combined with elastic network analysis to enhance global domain motions in atomistic molecular dynamics simulations.
  submitted to Proteins: Structure, Function, Bioinformatics

- Katja Ostermeir, Martin Zacharias.
  Rapid alchemical free energy calculation employing a Generalized Born implicit solvent model.
  submitted to the Journal of Physical Chemistry B

- Katja Ostermeir, Martin Zacharias
  Dynamic self- rearrangement of molecules in Molecular Dynamics simulations, to identify best binder
  in preparation

Publications not included in this thesis:

- Katja Ostermeir, Karen Alim, Erwin Frey
  Buckling of stiff polymer rings in weak spherical confinement
  Physical Review E, 81(6): 061802, 2010.

- Katja Ostermeir, Karen Alim, Erwin Frey
  Confinement induces conformational transition of semiflexible polymer rings to figure eight form
  Soft Matter, 6(15): 3467-3471, 2010.