# Computational Exploration of Protein Mechanisms at the Atomistic Scale

Maximilian Kienlein

Technische Universität München

TUM School of Natural Sciences

# Computational Exploration of Protein Mechanisms at the Atomistic Scale

Maximilian Kienlein

Vollständiger Abdruck der von der TUM School of Natural Sciences der Technischen Universität München zur Erlangung eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation.

| | |
|---|---|
| *Vorsitz:* | Prof. Dr. Markus Lackinger |
| *Prüfende der Dissertation:* | 1. Prof. Dr. Martin Zacharias<br>2. Prof. Dr. David Egger |

Die Dissertation wurde am 12.07.2024 bei der Technischen Universität München eingereicht und durch die TUM School of Natural Sciences am 26.09.2024 angenommen.

Für meine Eltern
Angela und Hermann

# Abstract

This thesis delves into the intricate world of protein mechanisms related to the cis/trans conformations of the amino acid proline (Pro) and the effect of ligand binding on structural transitions, utilizing advanced computational techniques.
The first part of this work focuses on the unique properties of Pro and its propensity for cis-isomerization in peptide bonds N-terminal to Pro. Overcoming the limitations of traditional molecular dynamics simulations, we introduce a novel computational scheme, $\omega$-bias potential replica exchange molecular dynamics ($\omega$BP-REMD), to accelerate cis/trans transitions. Application of $\omega$BP-REMD to Pro-containing tripeptides and gene-3 protein from phage fd demonstrates its superiority over standard Umbrella sampling. Furthermore, the method is capable to efficiently investigate correlated changes of isomerization states in proteins with multiple Pro residues. We extend the study of conformational coupling between the isomerization states of different Pro residues, using bradykinin (BK) as a model system. $\omega$BP-REMD allows comprehensive sampling and reveals the interdependence of individual Pro isomerization states. Surprisingly, the cis/trans equilibrium of a Pro residue is found to shift significantly depending on the isomerization state of other Pro residues, suggesting a crucial role for such coupling in larger folded proteins with additional conformational contraints.
The second part of this work shifts focus to substrate-binding domains (SBD) in the glutamine transporter GlnPQ, exploring the dynamics of the SBD2 domain. Molecular dynamics simulations unveil the conformational changes upon ligand binding and emphasize the intricate relationship between substrate binding and global domain closure. The study identifies a key interaction involving a C-terminal D1-tail$_{471-484}$ and a D2-helix$_{418-427}$, influencing the conformational switching barrier in SBD2. In the last chapter, the investigation extends to the role of SBD2 in nutrient uptake, addressing the inhibition of cellular glutamine uptake and SBD2 closure by L-arginine in solution, as observed in FRET experiments. Advanced simulations unveil the molecular details of arginine binding to SBD2, emphasizing Coulombic repulsion with Lys373 as a key factor hindering the closure of SBD2 and subsequent substrate translocation.

# Acknowledgements

I would like to say a big thank you to everyone who has supported me on my PhD journey. I'd especially like to thank my supervisor, Martin Zacharias, for introducing me to the fascinating world of computational biophysics. Martin, your patience and positive attitude have made a lasting impression on me. Your ever-present support and ability to find the silver lining in my challenges were always encouraging and uplifting. I'm grateful for that! I would like to thank Maria Reif for her support from the very first day as a PhD student. Maria, your constant help and willingness to invest your time and expertise has been invaluable. Thank you! I'd like to thank Sonja for keeping the ship afloat and especially for those morning coffee chats that started the day with a smile. To Till, Glenn, Korbi, Flo, PA and Paul, I express my gratitude for all their advice, shared sense of humour, and the memorable experiences we've shared within and beyond the confines of our good-old physics department. I'd like to thank Shu-Yu and Richard for all their help with biophysics and coding. They've also introduced me to the richness of Asian culture, including the finer nuances of Mandarin in the many enjoyable green tea sessions. A big thank you to the T38 crew, past and present, for creating a supportive and welcoming atmosphere for our biophysics endeavours! I'd especially like to thank Luis for all the epic chess battles and Christian, Brianda, Asha, Simone and Patrick!

And finally, I'd like to thank my family. My parents, for their unwavering support and for making this journey possible without ever placing expectations on me. And last but not least, my love Anka, thank you for always putting a smile on my face and reminding me of what is important in life.

# Contents

# 1 Introduction

> *Biology largely runs on the amazing tricks proteins can perform. But when it comes down to it, proteins are simply molecules obeying the laws of physics. We can think of them as machines, but there is no ghost inside. Proteins are completely inanimate objects, whose primary role is to fluctuate in conformation (i.e., in shape or structure). Biology, via evolution, has indeed selected for highly useful structural fluctuations, such as binding, locomotion, and catalysis. However, to understand these highly evolved functions in a given molecule or set of molecules, it is very informative to consider their spontaneous "wigglings and jigglings", to paraphrase Richard Feynman.*
>
> **— David M. Zuckerman**
> in "Statistical Physics of Biomolecules"

## 1.1 Proteins

Virtually every major process in the living cell is orchestrated by proteins, a fact aptly reflected in their name, originating from the Greek word "proteios", meaning "primary" or "standing in front" [1].
Hence, it might come as no surprise that, from a chemical point of view, proteins are by far the most structurally complex molecules known. Their intricate architectures derive from linear sequences comprised of 20 distinct amino acid building blocks, strung together like pearls on a necklace [2]. These fundamental components of the protein share a common scaffold but vary in their specific side chains, each possessing unique properties, as illustrated in Figure 1.1.

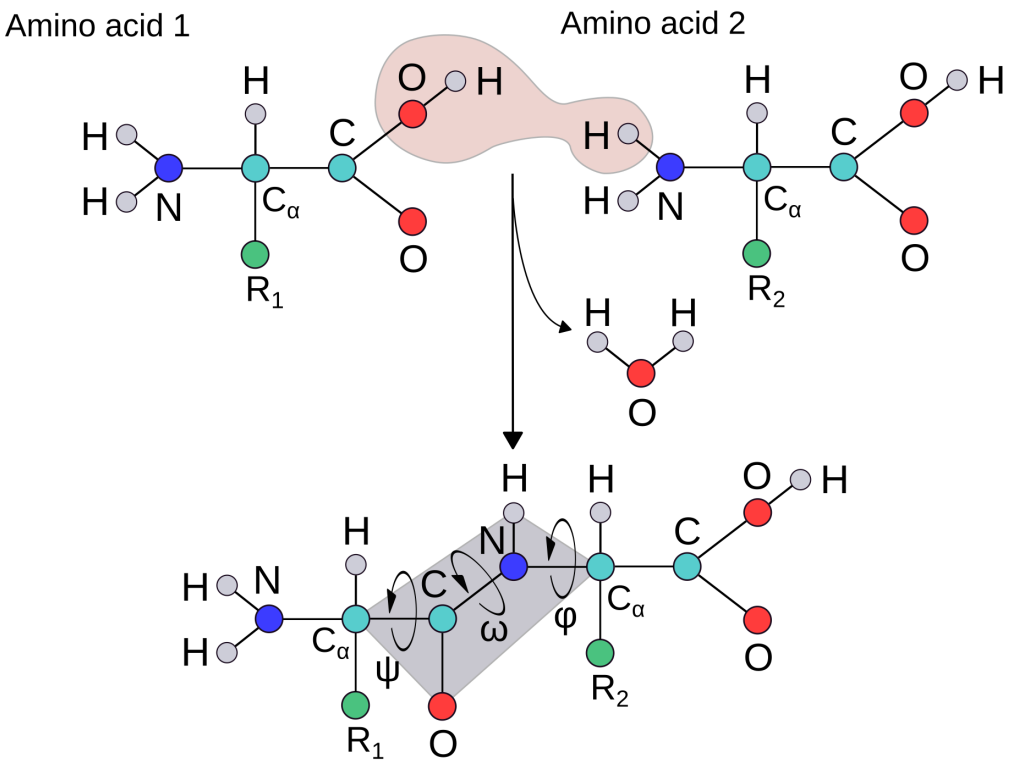



**Fig. 1.1.** Formation of a peptide bond between two amino acids in a condensation reaction, i.e. under expulsion of a water molecule. This process is repeated to form an extended chain that makes up a protein. The structural conformation of a protein is predominantly determined by the values of the $\psi$ (defined as the dihedral angle between N-$C_\alpha$-C-N atoms) and the $\phi$ dihedral angle (C-N-$C_\alpha$-C atoms). The $\omega$ dihedral angle ($C_\alpha$-C-N-$C_\alpha$ atoms) is (almost always) fixed at 180° to form the stable peptide plane, depicted in grey. The side chains ($R_1$ and $R_2$) attached to the $C_\alpha$ carbon atoms are unique to each amino acid.

The versatility encoded in this sequence of amino acids, known as the primary structure of the protein, is astounding. Proteins span a vast spectrum, from smaller peptides to colossal constructs containing more than 30,000 amino acids [3]. They act as transporters and catalysts, driving biochemical reactions as enzymes, regulating metabolism, and facilitating cellular processes [4–7]. Moreover, proteins can function as molecular switches and contribute significantly to cellular signaling, orchestrating pathways crucial for growth, development, and responses to external stimuli [8–10]. They form the foundation of cellular structures, providing support, shape, and movement [11, 12]. Without proteins, the intricate mechanisms within cells would come to a standstill and disrupt the delicate balance that is necessary for life-sustaining processes.

Proteins are assembled at the ribosome. In dependence on inherent physical interactions specific to their sequences and the surrounding environment, these initially elogated chains organize into elaborate three-dimensional structures [13, 14]. This complex process is known as protein folding [15]. The typical local spatial conformations these polypeptide chains fold into, such as $\beta$-sheets or $\alpha$-helices, constitute the secondary structure of the protein. The $\psi$ and $\phi$ dihedral angles (see Fig. 1.1) of the backbone scaffold typically arange into certain patterns when adopting these local folds, while the $\omega$ dihedral angle at the peptide bond remains fixed at 180°.
An exception to this rule is the amino acid proline, characterized by its 5-membered-ring side chain, which connects to the backbone at both the $C_\alpha$ carbon atom and the backbone nitrogen atom. This distinctive cyclic structure of the proline side chain imparts significant conformational rigidity compared to other amino acids. Notably, proline stands out as the only amino acid capable of readily adopting a cis-conformation ($\omega = 0°$) besides the otherwise strongly preferred trans conformation ($\omega = 180°$) – a distinctive feature conferring exceptional biological relevance to proline [16–18]. This fascinating phenomenon will be detailed in Chapters 3 and 4 of this thesis, dedicated to gain deeper understanding of this process through computer simulation.

The overall outcomes of protein folding are manifold, ranging from relative stable native folds to continuously interchangable states, or inherently disordered conformational ensembles [19–21]. These conformational states, however, are by no means up to the protein alone, as life, on the microscopic scale of a protein, operates within a thermodynamic ensemble where the extent of Brownian motion and thermal collisions significantly govern both structure and dynamics. As a

result, the atoms constituting a protein are in perpetual motion, driving continuous conformational fluctuations that render the protein devoid of a fixed or static structure. Thus, within the cellular milieu, defining a singular "structure" for a protein becomes an elusive concept. In fact, the functioning of a protein is generally dependent on structural fluctuations [22] and can only be understood within the framework of thermodynamics and statistical mechanics. The latter theoretical approach aims to connect an unfathomable multitude of microscopic states and interactions to predictable macroscopic quantities.

The fundamental premise that underlies this thesis is exactly this notion that all phenomena originate from the interactions among atomistic constituents. Computers serve as invaluable tools in realizing this approach quite literally: A model of the microscopic interactions in the system is implemented numerically on a computer to extract the "modes of motion" of its basic components and, through the machinery of statistical mechanics, enables predictions of macroscopic observables.

Of course, the degree of sophistication with which the system is described must be tailored to the specific problem at hand, balancing accuracy against computational affordability. Figure 1.2 illustrates this point schematically.

While quantum mechanics offers the most accurate depiction of atomistic interactions, by explicit treatment of nuclear and electronic degrees of freedom, its application to the majority of biomolecular systems is not feasible due to the immense computational costs involved.

Atomistic molecular dynamics simulations (MD) provide an alternative, enabling the study of molecular motions in larger systems over longer timescales within a classical Newtonian framework. Usually, the considered degrees of freedom are individual atoms.

Over the years, modern MD has made substantial strides, now complementing biomolecular experiments by simulating systems encompassing tens of thousands to millions of atoms for timescales ranging from picoseconds to microseconds [26, 27]. Advanced sampling techniques[28] have further extended the scope, allowing the examination of processes occurring on timescales beyond the reach of brute-force simulations. Such techniques make use of the power of statistical mechanics to drastically reduce the required timescales for the calculation of equilibrium properties. The capability of molecular dynamics (MD) simulations to replicate biomolecular processes at the atomistic level and achieve a femtosecond time resolution, which exceeds the capabilities of current experimental techniques, has earned MD the name "Computational microscope" [29].

Next, I will briefly outline the basic structure of this work and the systems upon which I focused this microscope.

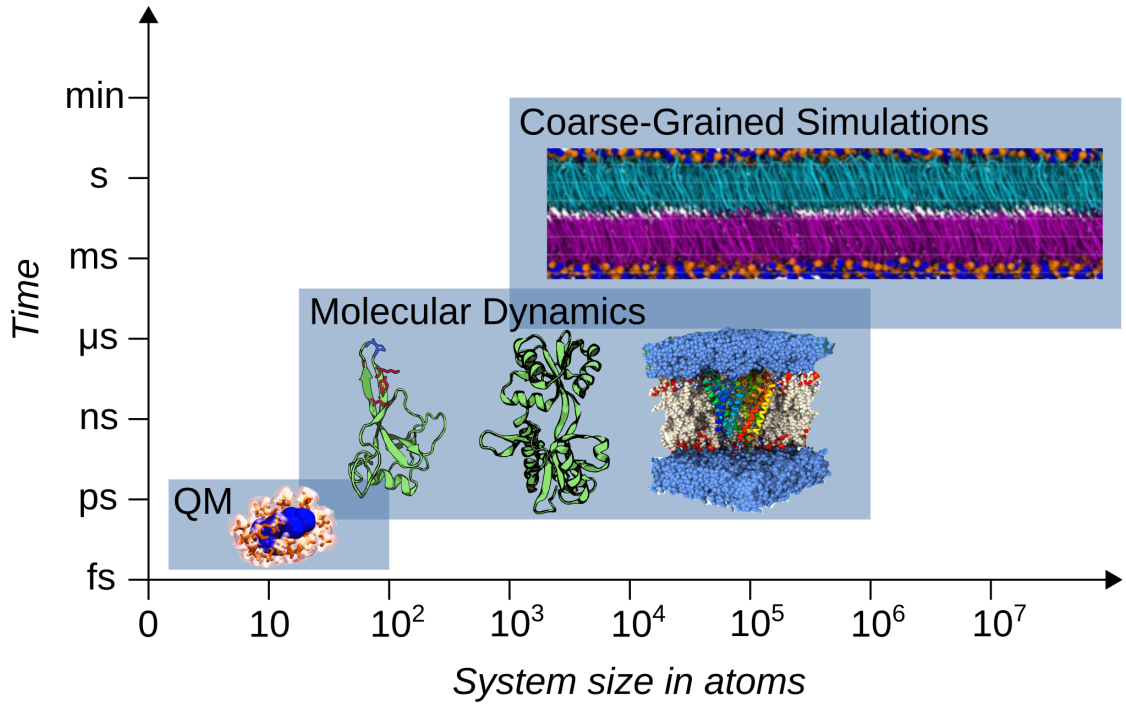



**Fig. 1.2.** Accessible timescales and system sizes of different biomolecular simulation techniques. Ab-initio molecular dynamics simulations offer precise treatment of both nucleic and electronic degrees of freedom by solving the Schrödinger equation. Due to immense computational costs, these simulations are limited to very small systems. Molecular dynamics (MD) simulations are based on a classical framework at the atomic resolution. Coarse-grained models dismiss atomic resolution to handle even larger systems, however, at a substantial cost of accuracy. Considering the sizes and relevant timescales of the biological system examined in this thesis, MD simulations prove to be the most suitable choice. Parts of this illustration have been modified from [23–25].

## 1.2 Thesis Outline

In the initial sections of my thesis, I lay out the theoretical groundwork, starting with fundamental principles of statistical mechanics, followed by an introduction to Molecular Dynamics (MD) simulations and related techniques.

In Chapter 3, a new method for advanced sampling, $\omega$BP-REMD, is introduced. This method was designed to facilitate the calculation of cis/trans isomerization free energies of the amino acid proline (Pro). Following a comprehensive description of the method, including the underlying theory, it is applied to investigate various Pro-containing tripeptides and a biologically significant Pro residue in the N2 domain of the gene-3-protein of phage fd. This chapter closely aligns with the findings and content published in: Kienlein, M.; Zacharias, M.; Reif, M. M. Efficient and accurate calculation of proline cis/trans isomerization free energies from Hamiltonian replica exchange molecular dynamics simulations. Structure 2023, 31(11), 1473-1484.

In Chapter 4, the examination continues by utilizing $\omega$BP-REMD to investigate the coupled cis/trans isomerization of the three proline (Pro) residues within the nonapeptide hormone bradykinin. This study delves into the conformational interdependence among various isomerization states of Pro residues – an aspect that is not well-understood. It underscores a significant strength of $\omega$BP-REMD in efficiently calculating higher-dimensional proline cis/trans isomerization events. This chapter has been published in similar form in: Kienlein, M.; Zacharias, M.; Reif, M. M. Comprehensive analysis of coupled proline cis-trans states in bradykinin using $\omega$BP-REMD simulations. JCTC 2024, 20(6), 2643-2654.

Chapter 5 shifts focus away from proline isomerization, directing our computational microscope towards another system – the substrate binding domain 2 (SBD2) of the GlnPQ transporter, an integral component for cellular glutamine uptake. Through extensive molecular dynamics (MD) simulations, we gained insights into both the ligand binding process and the global conformational rearrangements of SBD2 upon ligand binding. Apart from minor adjustments, this chapter was originally published in: Kienlein, M.; Zacharias, M. Ligand binding and global adaptation of the GlnPQ substrate binding domain 2 revealed by molecular dynamics simulations. Protein Science 2020, 29(12), 2482–2494.

In Chapter 6, the exploration of SBD2 is extended by addressing a question raised by FRET experiments. These experiments demonstrated that arginine in solution inhibits the cellular uptake of glutamine and prevents the closing of SBD2. Through the application of advanced sampling methods, we were able to shed light on the

underlying mechanisms of this phenomenon. The contents of this chapter have been published in similar form in: Kienlein, M.; Zacharias, M. How arginine inhibits substrate-binding domain 2 elucidated using molecular dynamics simulations. Protein Science 2024, 33(7), e5077

# 2 Theoretical Background

> *But thought's the slave of life, and life's time's fool,*
> *And Time, that takes survey of all the world,*
> *Must have a stop.*
>
> **— Shakespeare**

In this chapter, the fundamental equations of statistical mechanics are introduced. Subsequently, a brief overview of molecular dynamics (MD) simulations is provided, along with an overview of the computational techniques employed in this thesis to extract free energies from such simulations. Many books have been written on these methodologies and it would be hardly reasonable to cover them in the scope of a single chapter. For a thorough introduction to these subjects, please consult the cited literature references.

## 2.1 Statistical Mechanics - Basic Principles

Statistical mechanics provides the connection between macroscopic (measurable) quantities of many-body systems, e.g. temperature, pressure or volume and the underlying microscopic states. In statistical mechanics, the microscopic state of a system with $N$ particles is fully described by its location $\mathbf{x} = (q_1, ..., q_{3N}, p_1, ..., p_{3N})$ in phase space, where $p_i$ and $q_i$ denote the generalized coordinates and conjugate momenta, respectively. For a $N$-particle system the phase space is $6N$-dimensional. Each point in phase space evolves in time according to its Hamiltonian $H(\mathbf{x})$ [30]:

$$\frac{\partial H}{\partial p_i} = \frac{dq_i}{dt} \tag{2.1}$$

$$\frac{\partial H}{\partial q_i} = -\frac{dp_i}{dt} \tag{2.2}$$

The right-hand side of Hamilton's equations (2.1 and 2.2) assigns a "velocity" vector $\dot{\mathbf{x}}$ to each point in phase space. Thus, the time-evolution or trajectory of a system can be conceptualized as a flow in the "phase space fluid".
In statistical physics, one does not consider individual systems, but collections of systems sharing identical macroscopic properties. These collections are referred to as ensembles. The distribution of microstates of the ensemble in phase space at time $t$ is characterized by the phase space distribution function, denoted as $\rho(\mathbf{x}, t)$. It is important to recognize that the quantity $\rho(\mathbf{x}, t)d\mathbf{x}$ is the fraction of the entire ensemble members within the phase space volume $d\mathbf{x}$ at time $t$. Consequently, this expression is a probability density. Due to the incompressibility of the phase space fluid, the density $\rho$ remains conserved along any trajectory in the phase space flow, as stated by the Gibbs-Liouville equation [31–33]:

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \Big\{\rho, H\Big\} = 0, \tag{2.3}$$

where $\Big\{...\Big\}$ denote Poisson brackets. In thermodynamic equilibrium the macroscopic observables of an ensemble do not change over time. This is only possible, if the ensemble distribution has no explicit time dependence, implying that $\frac{\partial \rho}{\partial t} = 0$. As a consequence, Equation 2.3 simplifies to $\Big\{\rho, H\Big\} = 0$, allowing a general solution for the probability density $\rho$ to be any function $F$ of $H$ [34]:

$$\rho(\mathbf{x}) = \frac{1}{Z}\, F(H(\mathbf{x})) \tag{2.4}$$

with a central quantity in statistical mechanics, the partiton function $Z$

$$Z = \int d\mathbf{x}\, F(H(\mathbf{x})). \tag{2.5}$$

$Z$ serves to normalize $F$ and is a measure of the number of microscopic states accessible to a given ensemble. This way the link between the microscopic and macroscopic worlds is established. Macroscopic equilibrium quantities $A$ can be obtained as an average over the ensemble:

$$A = \langle a \rangle = \frac{1}{Z} \int d\mathbf{x}\, a(\mathbf{x}) F(H(\mathbf{x})) \tag{2.6}$$

If a single system over long periods of time was able to visit the whole phase space allowed by its Hamiltonian, it is referred to as being ergodic. In such instances,

the ergodic hypothesis states that the ensemble average in 2.6 can be equivalently expressed as a time average[35]:

$$A = \langle a \rangle = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau dt \; a(\mathbf{x}(t)) \tag{2.7}$$

To find specific solutions for $\rho(\mathbf{x})$, one must consider particular ensembles.
In the microcanonical ensemble the set of common macroscopic properties are the number of particles ($N$), volume ($V$) and energy ($E$). The definition of a specific energy value restricts the accessible phase space, permitting only those microstates for which $H(q_1, ..., q_{3N}, p_1, ..., p_{3N}) = E$. This restriction can be expressed in terms of the $\delta$ function, resulting in the following uniform microcanonical distribution function [36]:

$$\rho(\mathbf{x}) = \frac{\delta(E - H(\mathbf{x}))}{\Omega(E)} \tag{2.8}$$

with $\Omega$, the microcanonical partition function:

$$\Omega(N, V, E) = \frac{\Delta E}{h^{3N} N!} \int d\mathbf{p} \int_{S(V)} d\mathbf{q} \; \delta(E - H(\mathbf{q}, \mathbf{p})) \tag{2.9}$$

Here, $h$ denotes Planck's constant and $S$ is the spatial region the positions $q_i$ are restricted to when keeping the volume $V$ of the system constant. The term $N!$ accounts for the indistinguishability of particles[37], while $h^{3N}$ is the volume of a phase space cell. The size of these phase space volumes determine the resolution by which individual microstates can be distinguished from one another and is naturally set by Heisenberg's uncertainty relation [31]. When the phase space is partitioned into cells of finite size, the energy of a microstate cannot be precisely defined. Instead, the accessible phase space is confined to a thin energy shell with thickness $\Delta E$ above the constant energy hypersurface. Moreover, binning allows to exactly count the number of microstates available to the system. The corresponding thermodynamic potential, the entropy $S$, can be calculated from $\Omega$($N$,$V$,$E$):

$$S(N, V, E) = k_B \ln(\Omega(N, V, E)) \tag{2.10}$$

where $k_B$ denotes the Boltzmann constant.
However, the condition of constant total energy is hardly feasible under experimental conditions. Experimental setups typically utilize the temperature $T$ as a thermodynamic control variable. In such cases, other ensembles become relevant, such as the canonical ensemble ($N$,$V$,$T$) or the isothermal-isobaric ensemble ($N$,$P$,$T$). For the purpose of this discussion, the canonical ensemble will be considered. For larger systems, the canonical distribution serves as a reliable approximation for

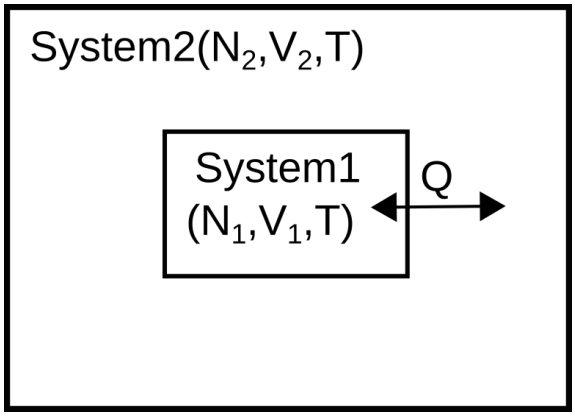



**Fig. 2.1.** Schematic representation of the canonical ensemble: System 1 is in contact with a thermal reservoir (System 2), allowing the exchange of heat $Q$. System 1 comprises $N_1$ particles in the volume $V_1$. System 2 comprises $N_2$ particles in the volume $V_2$. The thermal reservoir is considered to be sufficiently larger than system 1: System2 $>>>$ System1

the isothermal-isobaric or grand canonical ensemble ($\mu$,$V$,$T$), with the ensembles converging towards equivalence in the thermodynamic limit (on the order of $\mathcal{O}^{23}$ particles) [38]. Furthermore, the canonical expressions provide the groundwork for both the ($N$,$P$,$T$) and ($\mu$,$V$,$T$) ensembles [34]. One way to derive the expressions for the canonical distribution function and the associated partition function is to split up a microcanonical system into two parts and allow the exchange of heat. The surrounding system, often referred to as the thermal reservoir (System 2), is considered to be significantly larger than System 1. Consequently, the Hamiltonian of the smaller system is no longer conserved, as heat exchange can occur between the two systems. Nonetheless, in a canonical ensemble, the probability of occupying states diminishes exponentially with higher energy, as expressed by the famous Boltzmann factor[31]:

$$\rho(\mathbf{x}) \propto e^{-\beta H(\mathbf{x})} \tag{2.11}$$

with $\beta = 1/k_B T$. The corresponding partition function $Z$ for the canonical ensemble can be written as:

$$Z(N,V,T) = \frac{1}{h^{3N} N!} \int d\mathbf{x}\, e^{-\beta H(\mathbf{x})} \tag{2.12}$$

The free energy, denoted as $F = -k_B T \cdot \ln(Z)$, provides the corresponding potential from which all thermodynamic quantities can be derived.

As a final remark, it is important to note that in practical applications, none of the aforementioned partition functions are directly accessible due to the high-dimensional nature of the integrals in 2.6. Instead, in MD simulation, a single system is propagated in time in discrete steps, generating configurations in ac-

cordance with the underlying probability density. This way, the dynamics of the system can be reproduced and, under the assumption of ergodicity, the equilibrium quantities in Eq. 2.7 can be extracted from the simulation:

$$A = \langle a \rangle = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau dt \; a(\mathbf{x}(t)) = \frac{1}{N} \sum_{n=1}^{N} a(\mathbf{x}_{n\Delta t}) \tag{2.13}$$

Here, $a(\mathbf{x}_{n\Delta t}$ denotes values of the observable $a$ sampled at phase space positions that were accessed at discrete time steps $n\Delta t$.

## 2.2 Molecular Dynamics Simulation

After introducing the key concepts in statistical mechanics, we delve into more specific concepts related to Molecular Dynamics (MD) simulation in this section.

### 2.2.1 Basic Assumptions

MD simulations resort to a classical framework primarily due to computational limitations. The nuclei are treated as classical point particles with fixed partial charges, in contrast to expensive quantum mechanical (QM) calculations[39, 40]. In QM simulations, the electronic wave function is explicitly determined by calculating approximate algebraic solutions to the Schrödinger equation[41]. In life science ab-initio studies, the practical scope is restricted to smaller molecules. These simulations typically focus on system sizes in the order of $10^1$ atoms, studied in single-point calculations [42–48].

Classical treatment, however, can be justified not solely by a lack of alternatives in simulating biologically interesting systems, but also substantiated by several other reasons.
Firstly, the considerable mass disparities between nuclei and electrons allow for the separation of their motions as described by the Born-Oppenheimer approximation [49, 50]. Therefore, the rapid electronic movements instantaneously adjust to the slower nucleic motions and can be effectively represented as averages within classical approximations [51].
Secondly, justifying the classical approximation involves considering the thermal de Broglie wavelength [52]:

$$\Lambda = \sqrt{\frac{h^2}{2\pi m k_B T}} \tag{2.14}$$

Quantum effects are expected to become significant when $\Lambda$ is larger than the inter-particle distance. For heavy atoms in a protein at T =300 K, $\Lambda$ is smaller than a typical bondlength of $1$ Å (e.g. $\Lambda_{Oxygen} = 0.25$ Å, $\Lambda_{Nitrogen} = 0.27$ Å and $\Lambda_{Carbon} = 0.29$ Å). Note that hydrogen atoms do not satisfy the de Broglie criterion, given their thermal de Broglie wavelength $\Lambda_{Hydrogen}$ of $1$ Å. Nonetheless, it is presumed that the vibrations of hydrogen atoms have a negligible influence on the conformational changes observed in proteins, as these changes occur on significantly longer time scales. With these assumptions in place, numerous studies demonstrate that the dynamics of many biological systems can be effectively described by a classical formulation [53–58].

### 2.2.2 Force Field

The MD studies conducted in this thesis rely on a classical Hamiltonian, where atoms are treated as point particles with fixed partial charges. The corresponding potential energy function is commonly referred to as the "Force Field". This section provides an explanation of the fundamental terms associated with biomolecular force fields. Further details on the efficient computation of the long-range interaction terms will be provided in the upcoming section.

While maintaining simplicity in the functional expressions for computational efficiency is desirable, the refinement of force fields for biomolecular systems remains an ongoing endeavor. Specifically, the accurate parametrization of modern force fields is a highly non-trivial task and increasingly relies on extensive quantum mechanical calculations [59, 60] or machine learning approaches [61].
Despite these complexities, force fields typically adhere to the following functional form[62]:

$$\begin{aligned} U(\mathbf{r}_1, ..., \mathbf{r}_N) = & \sum_{Bonds} k_d (d - d_0)^2 + \\ & \sum_{Angles} k_\Theta (\Theta - \Theta_0)^2 + \\ \sum_{Dihedrals} & \frac{V_n}{2} \cdot [1 + \cos(n\Phi - \gamma)] + \\ & \sum_{i<j} \left[ \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\epsilon r_{ij}} \right] \end{aligned} \tag{2.15}$$

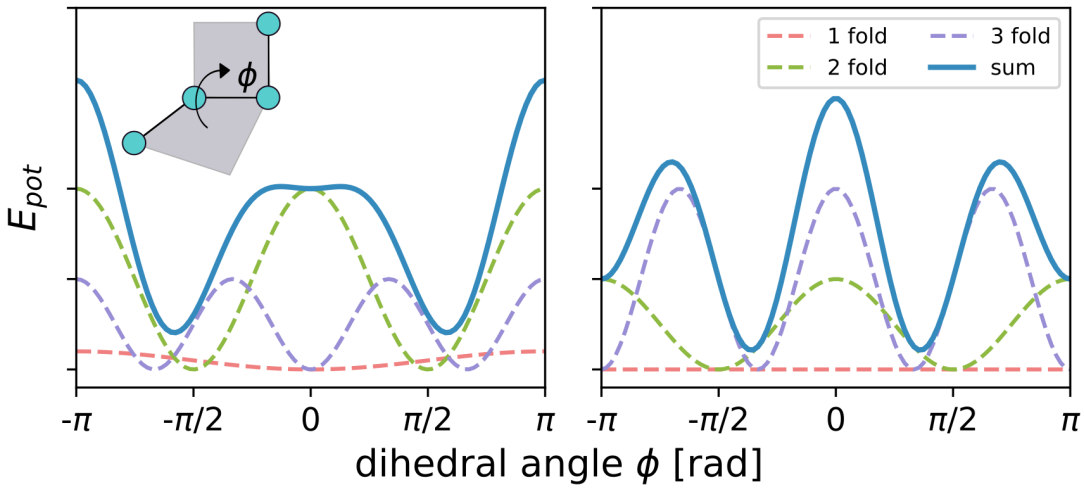



**Fig. 2.2.** Potential energy about a bond torsion as sum of a truncated Fourier series. The dihedral angle $\phi$ is defined by a group of four atoms. The truncated Fourier series, typically at third order, is capable to reproduce various energy landscapes with multiple minima, for example to adequately predict the torsional energy at cis/trans/gauche conformations.

The first two terms characterize the stretching and bending of covalently bound atoms in a harmonic potential form. The force constants, denoted as $k_d$ for stretching and $k_\Theta$ for bending, along with the equilibrium length $d_0$ and equilibrium angle $\Theta_0$, are typically empirically determined to match measured molecular vibrational spectra of smaller molecules [63].
The barriers to internal rotation about specific bonds is characterized by a Fourier series (illustrated in Figure 2.2), where $V_n$ represents the barrier heights, and the variables $n$ and $\gamma$ signify the multiplicities and phases, respectively. The physical origins of this term are complex and not completely clarified yet [64]. It is hypothesized that steric strain resulting from overlapping bond orbitals, resonance stabilization, and other quantum-mechanical effects play a significant role. Torsional parameters are typically determined through ab-initio quantum mechanical calculations of smaller model compounds, combined with geometry optimizations [65]. These parameters hold substantial importance in the formation of secondary structures, and most refinements of force fields revolve around optimizing these terms.

The last term in the summation accounts for non-bonded interactions. Lennard-Jones potentials, characterized by parameters $A$ and $B$, are utilized to describe van der Waals interactions between atom $i$ and atom $j$ at an interatomic distance $r_{ij}$[66]. $A$ and $B$ are specific to the atom types of the interacting atoms. A short-range, repulsive $1/r^{-12}$ term is incorporated to model the Pauli exclusion principle

[67], while the attractive $1/r^{-6}$ term originates from quantum mechanics, reflecting electron correlation contributions, commonly known as London or dispersion forces[68]. Parameters for the van der Waals potential can be derived by fitting parameters to lattice energies and crystal structures, or to reproduce observed liquid properties, like e.g. solvation free energies, heat capacities and densities in simulations.
Finally, the computation of long-range electrostatic interactions, according to Coulomb's law, relies on the partial charges $q_i$ and $q_j$ of the interacting atoms.

It is important to note that the parameters used in this additive potential function are specific to individual atom types and combinations of atom types. In practice, multiple atom types are defined for atoms of the same element to account for different chemical environments. As a consequence, this leads to a relatively extensive set of parameters with numerous degrees of freedom to construct a empirical energy surface. Currently, there exists a wide array of force fields, each optimized for specific applications [69–71].

### 2.2.3 Periodic Boundary Conditions and Electrostatic Interactions

To accurately replicate physiological conditions, MD simulations must incorporate solvent molecules such as water and ions that surround the system of interest. One computationally efficient approach, albeit less precise, involves implicitly representing these molecules as a dielectricum [72].

In this thesis, however, water and ion molecules were explicitly simulated. This poses challenges when attempting to confine these atoms within a simulation box. Note, that for typical system sizes in the range of $10^5$ atoms, a significant fraction of atoms is expected to be located near the box boundaries ($\sim$ 12 % for a cubic box assuming a crystal-like configuration). This proximity can lead to serious artifacts [73]. To address this issue, we employ periodic boundary conditions (PBC), allowing emulation of a considerable number of solvent molecules without constraining them within finite boundaries.

With PBC, space is divided into identical cuboids (or truncated octahedrons), treating them as equivalent representations of the central simulation box. As illustrated

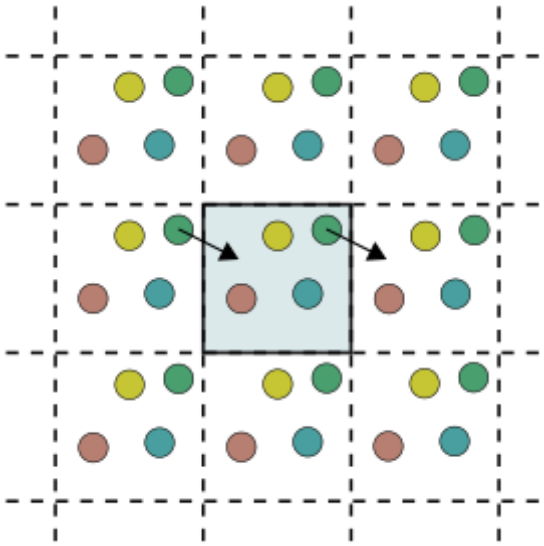


**Fig. 2.3.** Periodic boundary conditions in 2D. The unit cell in the center is surrounded by periodic images. A particle leaving the unit cell is re-imaged back into the central box.

in Fig. 2.3, when particles exit the central box, their coordinates are remapped to the opposite side. However, the replication introduced by periodic boundary conditions poses significant challenges in computing the non-bonded force field terms outlined in Eq. 2.15. Pairwise calculations alone scale with $\mathcal{O}(N^2)$ for an $N$-atom system, becoming prohibitively expensive for typical system sizes ranging from $N \sim 10^3$-$10^5$ atoms within the unit cell. Moreover, with periodic boundary conditions, the number of pairwise interactions to be calculated theoretically becomes infinite.

To manage this issue, simple spherical cut-off schemes (typically applied at $\sim 9$ Å) effectively handle the rapidly decaying Lennard-Jones potential ($\sim r^{-6}$). However, these schemes introduce significant inaccuracies when dealing with the more slowly decaying Coulombic potential [74]. In practice, electrostatic interactions are treated with the Ewald summation method [75]. This approach splits the electrostatic potential into a short-range part, converging quickly in real space, and a smooth long-range part, that can be represented by a rapidly converging Fourier series. In practice, this is achieved by extending the direct pairwise sum with contributions from Gaussian screening densities according to Fig. 2.4.

Setting the net charge of the unit cell to zero and neglecting a dipolar term, the potential can be expressed as [76]:

$$U_{Coulomb} = U_{short-range} + U_{long-range} + U_{self} \tag{2.16}$$

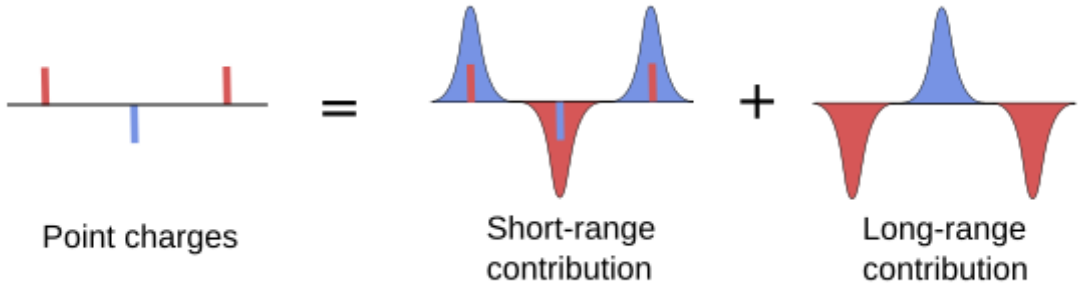



**Fig. 2.4.** In the Ewald summation method a set of point charges is considered as a sum of screened charges and the corresponding neutralizing charge densities.

The solutions to the respective Poisson equation lead to the following expression for each of the contributions:

$$U_{short-range} = \frac{1}{2}\sum_{i\neq j}^{N} q_i q_j \,\mathrm{erfc}(\sqrt{\alpha} r_{ij})/r_{ij} \tag{2.17}$$

where $\mathrm{erfc}(x)$ is the complementary error function and $\alpha$ the inverse width of the Gaussian screening function. The complementary error function exponentially decays with distance, allowing truncation of $U_{short-range}$ in real space at a specified cut-off radius ($\sim 9$ Å, equivalent to the van der Waals terms).

The long-range contributions to the Coulomb potential, arising from periodically arranged Gaussian charge distributions, are effectively calculated in Fourier space.

$$U_{long-range} = \frac{1}{2V}\sum_{\mathbf{k}\neq 0} \frac{4\pi}{k^2} |\rho(\mathbf{k})|^2 \exp(-k^2/4\alpha) \tag{2.18}$$

where $\rho(\mathbf{k})$ is the structure factor. In reciprocal space, these contributions are short-ranged, as the sum rapidly decays with larger wave numbers $|\mathbf{k}|$ and can be truncated at a cut-off wave number [73].
Finally, a self-interaction term from the calculations of $U_{long-range}$ needs to be substracted to remove overcounting the interaction of a point charge $q_i$ with the centre of its Gaussian.

$$U_{self} = -(\alpha/\pi)^{1/2} \sum_{i=1}^{N} q_i^2 \tag{2.19}$$

Further speed up can be achieved by the Particle Mesh Ewald method [77]. Here, the charge posititons get interpolated to a grid which allows the application of Fast Fourier Transformation (FFT). Through this process, a reduction in the computa-

tional complexity of calculating long-range interactions is achieved, decreasing from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$.

### 2.2.4 Integrating the Equations of Motion

After establishing the model used in MD simulation, a numerical solver is required to propagate the system through time in accordance with Hamilton's equations of motion (2.1 and 2.2). Numerous finite difference methods are available for this purpose, such as the leap-frog algorithm, Verlet algorithm, or velocity Verlet algorithm, among others [78–80]. In the following the velocity Verlet-algorithm is introduced,which explicitly updates positions and velocities. This approach elegantly aligns with the phase space picture, taking into account both positions and momenta (or velocities). A step forward in time by an amount $\delta t$ can be readily obtained through Taylor series expansion:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \mathbf{v}_i(t) + \frac{\Delta t^2}{2m}\mathbf{F}_i(\mathbf{r}(t)) + \mathcal{O}(\Delta t^3) \tag{2.20}$$

where $\mathbf{r}_i$ and $\mathbf{v}_i$ are positions and velocities of particle $i$ and $\mathbf{F}_i$ the force on this particle, while $\mathbf{r}$ denotes the 3N-dimensional position vector of all $N$ particles. An expression for the velocity updates can be derived by consideration of a step backwards in time:

$$\mathbf{r}_i(t) = \mathbf{r}_i(t + \Delta t) - \Delta t \mathbf{v}_i(t + \Delta t) + \frac{\Delta t^2}{2m}\mathbf{F}_i(\mathbf{r}(t + \Delta t)) - \mathcal{O}(\Delta t^3) \tag{2.21}$$

Inserting 2.20 into 2.21 and solving for $\mathbf{v}_i(t + \Delta t)$ yields:

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \frac{\Delta t}{2m_i}\left[\mathbf{F}_i(\mathbf{r}(t)) + \mathbf{F}_i(\mathbf{r}(t + \Delta t))\right] \tag{2.22}$$

Forces can be extracted from the force field potential via:

$$\mathbf{F}_i(\mathbf{r}) = -\nabla_i U(\mathbf{r}) \tag{2.23}$$

Important properties of the verlocity Verlet alogorithm are time-reversibility and symplecticity[81]. The timestep size $\Delta t$ has to be chosen according to the fastest motions of the system. Vibrations of the light hydrogen atom bonds would in principle require a timestep of $\Delta t \sim 0.5 - 1.0$fs. These rapid fluctuations are of quantum-mechanical nature and not well described by a harmonic potential. In addition, they are of minor biological relevance for long-term conformational changes

and are commonly constrained by Langrange multiplier methods[82]. Combined with other methods [83] to supress the fast hydrogen vibrations a timestep of $\Delta t = 4$fs becomes feasible.
So far, the introduced integration schemes conserve the Hamiltonin of the system, which implies the microcanonical ($N$,$V$,$E$) ensemble. Other ensembles like the canonical ($N$,$V$,$T$) or the isobaric ($N$,$P$,$T$) ensemble can be simulated by coupling to thermostats [84–87] or barostats [85], respectively.
To initiate a simuation, the starting positions are usually derived from high-resolution experiments like nuclear magentic resonace spectroscopy (NMR) or X-ray crystallography. The initial velocities can be assigned for example according to a Maxwell-Boltzmann distribution.

### 2.2.5 Advanced Sampling: Hamiltonian Replica Exchange

Propagating a system in phase space by simply integrating the equations of motion,i.e. performing local steps according to a physical force-field description as covered above, is commonly referred to as "free" or "brute-force" MD simulation. Typically, with this brute-force approach, only a limited region of phase space is explored because rough energy landscapes are sampled very inefficiently. Consider, for example, the Arrhenius equation, which provides a first-order approximation for the transition rate from state A to state B, separated by an activation free energy barrier of $\Delta G^{\ddagger}_{A\to B}$ [88]:

$$k_{A\to B} \propto e^{-\beta \Delta G^{\ddagger}_{A\to B}} \tag{2.24}$$

The probability of crossing a barrier in a given amount of time at a certain temperature decreases exponentially with the barrier height. As a result, escaping local minima with sufficient depth can be a very time-consuming process. To ensure proper sampling of all important states in their correct proportion, it is essential to repeatedly traverse such barriers. A commonly employed technique to improve sampling is to run multiple MD simulations with different Hamiltonians in parallel, accompanied by Monte Carlo exchange trials. This strategy is known as Hamiltonian replica exchange molecular dynamics simulation (HREMD)[89].

In HREMD $N$ noninteracting replicas are run simultaneously with $N$ different Hamiltonians $H$ at the same temperature $T$. The state of the generalized ensemble $X$ can be expressed as

$$X = [H_1(q_1), ..., H_i(q_i), H_j(q_j), ..., H_N(q_N)] \tag{2.25}$$

with the respective conformations $q$ of the individual replicas. In order to reach a extended Boltzmann equilibrium the detailed balance condition must be satisfied for all exchange attempts:

$$p(X)w_{ij} = p(X')w_{ji} \tag{2.26}$$

here $p(X)$ and $p(X')$ denote the probabilities of a given generalized state before and after a successful exchange between replicas $i$ and $j$, that is

$$X' = [H_1(q_1), ..., H_i(q_j), H_j(q_i), ..., H_N(q_N)] \tag{2.27}$$

and $w$ being the respective transition probabilities. Since all replicas are independent, the joint probability distribution for the extended system can be written as multiplication over each of the $N$ replicas:

$$P(X) = \prod_{i=1}^{N} P_i(q_i, H_i) \tag{2.28}$$

With this the detailed balance equation 2.26 can be written as

$$\frac{w_{ij}}{w_{ji}} = \frac{e^{-\beta(H_i(q_j)+H_j(q_i))}}{e^{-\beta(H_i(q_i)+H_j(q_j))}} = e^{-\beta\Delta} \tag{2.29}$$

where $\Delta = H_i(q_j) + H_j(q_i) - H_i(q_i) - H_j(q_j)$ . Using the Metropolis criterion, the probability of accepting an exchange $X \to X'$ is given by [90]

$$P_{acc} = min(1, e^{-\beta\Delta}) \tag{2.30}$$

This way the replicas can explore phase space with different Hamiltonians and configurations with favourable energies can traverse through the replica ladder while still sampling from a Boltzmann distribution. In this thesis, HREMD methods have been employed to enhance the sampling of cis/trans isomerization events of proline (details are provided in chapter 3) and in conjunction with Umbrella sampling (see Section 2.3.2)[91].

## 2.3 Free Energy Calculation

In Section 2.1, we introduced the concept of (absolute) free energy as a quantity proportional to the logarithm of the partition function. This function allows us to derive various thermodynamic quantities through differentiation.
However, when dealing with biomolecular systems, the focus often shifts from the

absolute free energy of the entire system to the free energy differences between specific states of the system.
Defining the microstates that populate a particular biological state can be approached in numerous ways. The first part of this thesis, for example, focuses on the free-energy differences between state A, encompassing all configurations where a proline residue exists in the trans isomer, and state B, where the proline adopts the cis isomer. Another example involves defining state A as comprising all ligand-receptor bound microstates, while state B encompasses all unbound states. The Boltzmann-weighted free energy $F$, i.e. $e^{-\beta F}$ of a given state is directly proportional to the probability of finding the system in that particular state [92]. Free-energy differences between states provide valuable insights into whether a process, transitioning the system from state A to state B, can occur spontaneously or if work must be invested to induce such a transition [93]:

$$\frac{p(A)}{p(B)} = \frac{e^{-F_A/k_BT}}{e^{-F_B/k_BT}} \tag{2.31}$$

It is worth noting that by taking, for example, the Helmholtz free energy $F = E - TS$ and plugging it into the probability term,

$$p(A) \propto e^{-F_A/k_BT} = e^{+S_A/k_B} e^{-E_A/k_BT} \tag{2.32}$$

one can easily see that a state can acquire probability either by increasing the number of contributing microstates or by lowering its average energy.

Given adequate sampling, free-energy differences can be extracted from a MD simulation by simply measuring the time spent in the respective states A and B and applying the results into Eq. 2.31.
However, as discussed in the previous section, the diffusion over substantial energy barriers in free simulation is highly inefficient, rendering this direct approach rarely applicable. In HREMD simulations (see preceeding Section 2.2.5) phase-space exploration is more efficient. Here, the extensive sampling frequently allows for the utilization of Eq. 2.31.
In the upcoming sections, further methods used in this thesis to extract free energies from simulations, will be introduced.

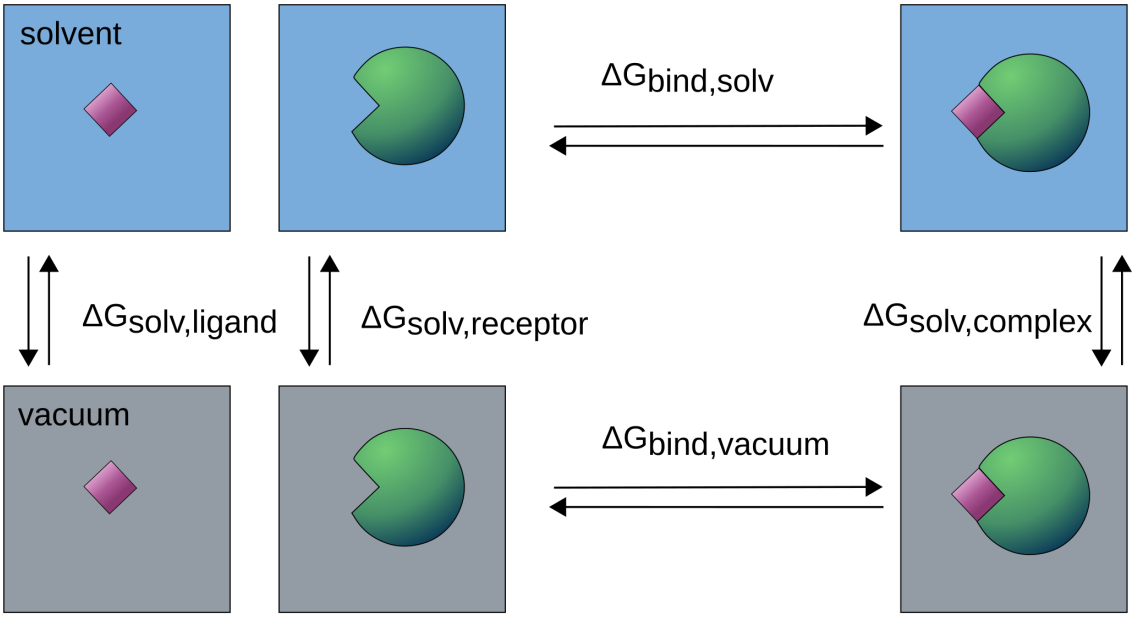



**Fig. 2.5.** The thermodynamic cycle illustrates the process of a ligand (violet) binding to a receptor (green) in a solution (blue). The calculation of the free-energy change, $\Delta G_{bind,solv}$, for this reaction is not directly computed due to the influence of perturbing solvent-solvent interactions. Instead the MMGBSA(/PBSA) approach splits this calculation up into different terms. $\Delta G_{solv,complex}$, $\Delta G_{solv,receptor}$, and $\Delta G_{solv,ligand}$ represent the solvation free energies of the three states calculated by either solving the Generalized Born or the Poission Boltzmann equation. Furthermore, $\Delta G_{bind,vacuum}$ can be calculated by taking into account the average interaction energy and the entropy change upon binding.

### 2.3.1 The MM-PBSA and MM-GBSA Approach

The molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) and its counterpart, the Generalized Born (MM-GBSA) approach, offer computationally inexpensive methods for post-processing trajectories to assess the free-energy difference between two states[94]. Typically, these states correspond to the bound and unbound configurations of two solvated molecules, as depicted in Fig. 2.5. In principle, $\Delta G_{bind,solv} = \Delta H - T\Delta S$ can be directly calculated from explicit solvent simulations of the bound an unbound state [95]. The huge amount of solvent-solvent interactions, however, leads to energetic fluctuations of the system which are orders of magnitude higher than the desired binding free energy resulting in poor convergence of such attempts.
In MM-PBSA/MM-GBSA methods the binding free energy is split up according to the thermodynamic cycle in Fig. 2.5:

$$\Delta G_{bind,solv} = \Delta G_{bind,vacuum} + \Delta G_{solv,complex} - (\Delta G_{solv,receptor} + \Delta G_{solv,ligand}) \quad (2.33)$$

$\Delta G_{bind,vacuum}$ can be determined by the calculation of the average interaction energy between receptor and ligand [96]. The conformational entropy contribution are usually determined by performing normal mode analysis [97] on the three species. In practice entropy contributions can be neglected if only a comparison of states of similar entropy is desired such as two ligands binding to the same protein. The reason for this is that normal mode analysis calculations are computationally expensive and tend to have a large margin of error that introduces significant uncertainty in the result.

$$\Delta G_{bind,vacuum} = \Delta H_{MM} - T\Delta S \tag{2.34}$$

The average interaction energies of receptor and ligand are computed by performing calculations over an ensemble of uncorrelated snapshots from an equilibrated molecular dynamics simulation. In principle, this approach would require three independent MD simulations (of the complex and of the protein and the ligand alone) in order to obtain the binding free energy. Here, the approximation that no significant conformational changes occur upon binding is made. Therefore the snapshots for all three species are taken from a single trajectory. This method is called "single trajectory approach" [98].

Solvation free energies entail electrostatic (Coulomb or polar) and hydrophobic (nonpolar) contributions

$$\Delta G_{solv} = (\Delta G_{coulomb,\epsilon=80} - \Delta G_{coulomb,\epsilon=1}) + \Delta G_{non-polar} \tag{2.35}$$

Electrostatics are calculated by solving the linearized Poisson Boltzmann equation

$$\nabla[\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})] = -4\pi\rho(\mathbf{r}) - 4\pi\lambda(\mathbf{r})\sum_i z_i c_i \exp(-z_i\phi(\mathbf{r})/k_B T) \tag{2.36}$$

where $\epsilon$ is the dielectric constant, $\phi$ the electrostatic potential, $\rho$ the charge of the molecule, $\lambda$ denotes the Stern function and $z_i$ the ion charge of a species with concentration $c_i$. From $\phi$ the solvation free energy can be calculated by summation over all charges $i$ of the solute:

$$\Delta G_{solv,polar} = \frac{1}{2}\sum_i q_i(\phi_i(\mathbf{r})) \tag{2.37}$$

In MMGBSA the Generalized Born equation can be used to approximate the PB equation[99]:

$$\Delta G_{polar} = -\frac{1}{2}\sum_{ij}\frac{q_i q_j}{f_{GB}}\left(\frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{out}}\right) \tag{2.38}$$

$$f_{GB} = \sqrt{r_{ij}^2 + R_j R_i e^{-D_{ij}}}, D_{ij} = r_{ij}^2 / 4 R_i R_j \tag{2.39}$$

Here, $f_{GB}$ is a smoothing function and R the effective Born radii for each atom $i$,$j$. The dielectric constant $\epsilon_{in}$ of the molecule is usually set to 1 and $\epsilon_{out}$= 80 for the solvent. For non-polar contributions an empirical term proportional to the solvent-accessible-surface-area (SASA) is added[100]:

$$\Delta G_{non-polar} = \gamma \cdot SASA \tag{2.40}$$

### 2.3.2 Reaction Coordinates and Umbrella Sampling

A key concept in molecular biophysics is to parameterize the progress of a certain process using a singular degree of freedom derived from a subset of generalized coordinates within the system. When generalized coordinates are used in this manner, they are commonly referred to as reaction coordinates or collective variables [34, 101]. Within the canonical ensemble, the probability density along the reaction coordinate $\xi(\mathbf{x})$ can be computed through a Boltzmann-weighted summation across all configurations that contribute to a specific value of $\xi$:

$$P(\xi) = \frac{1}{Z} \int d\mathbf{x}\, \delta(\xi(\mathbf{x}) - \xi) e^{-\beta H(\mathbf{x})} \tag{2.41}$$

The potential of mean force (PMF) $W(\xi)$, also known as the free energy profile along the $\xi$ coordinate, can be straightforwardly derived from this probability distribution:

$$W(\xi) = -k_B T \ln(P(\xi)) \tag{2.42}$$

Commonly used examples of reaction coordinates applied in computational biophysics involve metrics such as the distance between receptor and ligand, conformational changes characterized by radii of gyration or root-mean-square deviations, among many others. Moreover, the PMF enables insights into equilibrium distributions and transition barriers.

In principle, given that MD simulations sample in proportion to the Boltzmann factor, PMFs can be readily computed from a histogram constructed along the $\xi$ coordinate. Proper sampling of the reaction coordinate, as we already know, becomes challenging due to the inefficient diffusion across free energy barriers during free simulations. To address this, one approach involves introducing supplementary harmonic potentials that steer the system along the reaction coordinate. This methodology is commonly known as Umbrella sampling[102]. Umbrella sampling is usually performed in a large number of windows $i$, each having one biasing

potential $w_i(\xi) = 0.5k(\xi - \xi_i^{ref})^2$ with a predefined equilibrium position $\xi_i^{ref}$ and a force constant $k$. The modified distribution for each window is then given by:

$$\tilde{P}_i(\xi) = \frac{1}{\tilde{Z}} \int d\mathbf{x} \delta(\xi(\mathbf{x}) - \xi_i^{ref}) e^{-\beta(H(\mathbf{x}) + w_i(\xi))} \tag{2.43}$$

with the partition sum of the biased Hamiltonian for each window:

$$\tilde{Z} = \frac{1}{h^{3N} N!} \int d\mathbf{x} e^{-\beta H(\mathbf{x} + w_i(\xi(\mathbf{x})))} \tag{2.44}$$

Next, let us try to retain the unbiased probability distribution $P_i\xi$ for each window $i$:

$$\begin{aligned} \tilde{P}_i(\xi) &= \frac{1}{\tilde{Z}} \int d\mathbf{x} \delta(\xi(\mathbf{x}) - \xi) e^{-\beta(H(\mathbf{x}) + w_i(\xi))} \\ &= \frac{Z}{\tilde{Z}} e^{-w_i(\xi)} \frac{1}{Z} \int d\mathbf{x} \delta(\xi(\mathbf{x}) - \xi) e^{-\beta H(\mathbf{x})} \\ &= \frac{Z}{\tilde{Z}} e^{-w_i(\xi)} P_i(\xi) \end{aligned} \tag{2.45}$$

Thus, we arrive at:

$$P_i(\xi) = \frac{\tilde{Z}}{Z} e^{w_i(\xi)} \tilde{P}_i(\xi) \tag{2.46}$$

and the unbiased PMF along $\xi$ in the respective window can be recalculated from the biased probability distribution sampled in the umbrella simulations:

$$W_i(\xi) = -k_B T \ln(\tilde{P}_i(\xi)) - w_i(\xi) + F_i \tag{2.47}$$

where $F_i$ is a constant offset.
To combine all windows, the weighted histogram analysis method (WHAM) equations were used in this thesis [103]:

$$P(\xi) = \frac{\sum_{i=1}^{N} n_i P_i(\xi)}{\sum_{i=1}^{N} n_i e^{\beta F_i} e^{-\beta w_i(\xi)}} \tag{2.48}$$

$$e^{-\beta F_i} = \int d\xi P(\xi) e^{-\beta w_i(\xi)} \tag{2.49}$$

where $n_i$ is the number of sampled configurations in the $i$-th umbrella simulation. As both $P(\xi)$ and $F_i$ are unknown, these equations must be solved iteratively until self-consistency is achieved. A prerequisite for convergence is sufficient sampling overlap between neighboring umbrella windows [104].

# 3

# $\omega$BP-REMD: Proline Isomerization Unveiled by Molecular Dynamics Simulations [1]

> *Ihr könnt draußen in der Sonne abhängen und irgendwelche Bälle rumwerfen oder an euren Computern was Bedeutsames leisten.*
>
> **— Eric Cartman**

Proline (Pro) stands out among amino acids with its unique backbone side chain ring structure, which promotes the preferential adoption of cis isomers in peptide bonds N-terminal to Pro. Pro cis/trans isomerization is of major significance in numerous biological processes but typically occurs over extended timescales, ranging from seconds to minutes, posing challenges for brute-force MD simulations.
In this chapter we address this limitation by introducing a computational scheme, $\omega$-bias potential replica exchange molecular dynamics ($\omega$BP-REMD). This method involves modifying the Hamiltonian by reducing barrier heights in the energy profile of the peptide bond dihedral angles across a replica ladder. Additionally, an extra bias potential destabilizes a specific isomer state, ensuring balanced sampling between the two isomer states.
Application of $\omega$BP-REMD to various Pro-containing tripeptides and a crucial Pro residue in the N2 domain of gene-3-protein from phage fd, encompassing both wild-type and mutant protein variants, yielded notably accelerated cis/trans transition rates. In comparison, $\omega$BP-REMD outperforms standard Umbrella sampling (US) in terms of convergence and demonstrates superior agreement with experimental data. Notably, this method minimizes structural perturbations near the Pro residue and provides an efficient means to explore correlated changes in Pro isomerization

[1] This chapter has been published in similar form [105].

states within proteins. Exploring these correlated cis/trans isomerizations in the presence of multiple proline residues will be the focus of Chapter 4.

## 3.1 Introduction

Protein backbones are inherently characterized by a sequence of elements, including peptide bonds, with repeating $\phi$, $\psi$, and $\omega$ dihedral angles [2]. The $\phi$ and $\psi$ dihedral angles, corresponding to the torsions around the N-$C_\alpha$ and $C_\alpha$-C bonds, respectively, define the structure of the backbone in that particular values are characteristic of typical secondary structure elements. The $\omega$ dihedral angle corresponds to the torsion around the C-N bond and tends to be planar and rigid due to the delocalization of the $\pi$-electrons of the carbonyl group and the lone electron pair of the nitrogen atom, resulting in a resonance structure with a partial C-N double bond character. Steric repulsion between the side chains at consecutive $C_\alpha$ atoms in the cis state leads to an energetically more favorable trans isomer [106].
X-ray crystallography and NMR-based structural analysis of a vast range of folded proteins showed that essentially all $\omega$ dihedral angles are in the trans state with the exception of peptide bonds where the second (C-terminal) residue is a proline (Pro)[107].
Due to its cyclic structure (Fig. 3.1), Pro is the only amino acid showing less energetic discrimination between the cis and trans isomer states, i.e. the instability of the cis isomer is reduced due to steric similarity between the $C_\alpha$ and $C_\delta$ atoms. In fact, about $7\%$ of the peptide bonds preceding a Pro residue are present in the cis isomer [107]. The relatively comparable stability of cis and trans isomer states for Pro residues, as opposed to other amino acids, confers exceptional biological significance to (certain) Pro residues. The rotational energy barrier between the two isomers is about 20 kcal·mol$^{-1}$ at ambient conditions[108], which implies that the isomerization between the two conformational minima, referred to as cis/trans isomerization, is a very slow process. A cis/trans transition usually occurs on a time scale of 10-1000 s.
The capacity for cis/trans isomerization, coupled with the intrinsically slow isomerization mechanism, forms the foundation for the biologically significant functions of Pro residues in proteins. For example, they allow Pro residues to act as backbone switches involved in the regulation of biological processes ranging from protein folding [109], cell signaling cascades [110], ion channel gating [111] to neurodegeneration [112], gene expression [113] and more [114]. Dysfunctions in these control mechanisms have been related to various diseases such as the Alzheimers's

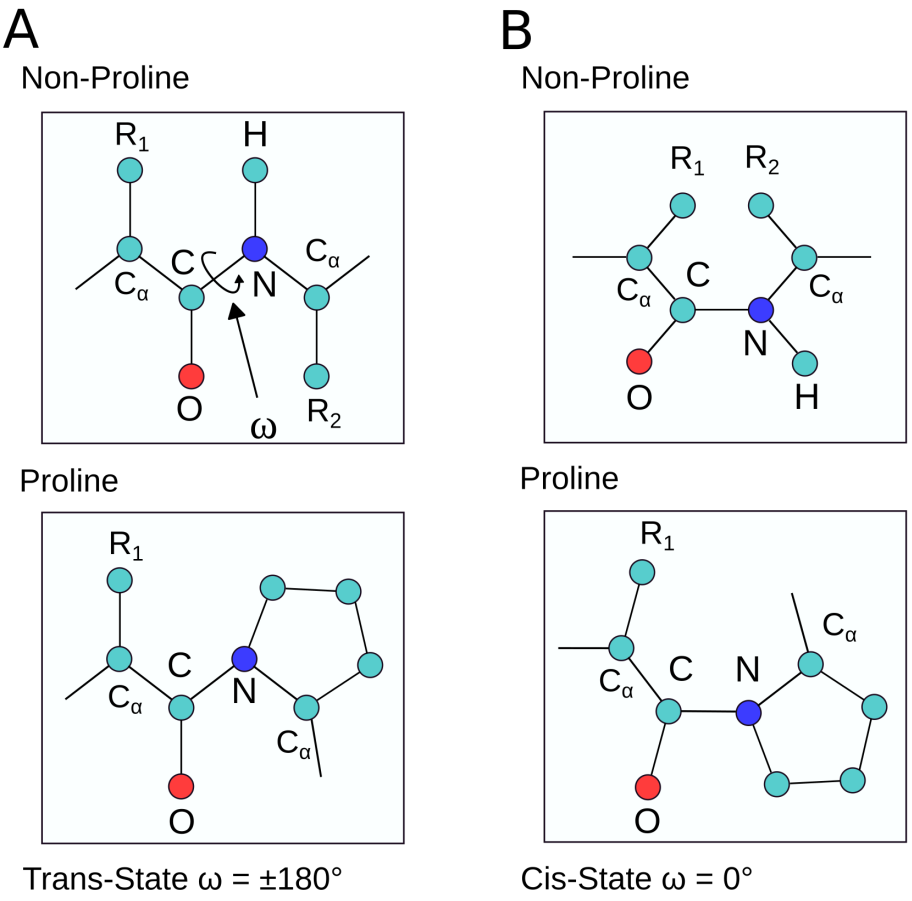



**Fig. 3.1.** Ball-and stick representation of a peptide bond between non-Pro amino acids (side chains $R_1$ and $R_2$) or a non-Pro (side chain $R_1$) and a C-terminal Pro residue in the (A) trans ($\omega = \pm 180°$) or (B) cis ($\omega = 0°$) isomer. Non-Pro peptide bonds are predominantly in the trans isomer due to steric clashes between adjacent side chains in the cis state. This imbalance is heavily reduced for Pro side chains since the nitrogen atom is attached to the $C_\delta$ carbon atom and both isomers experience similar steric clashes.

disease and cancer [115]. However, not all Pro residues are essential for the regulation of these processes, for example protein folding may continue at a normal rate despite having non-native isomer states of certain Pro peptide bonds [116].

Despite the potential significance of Pro residues in biological processes, it is not clear how to explain or possibly predict the functional importance or functional irrelevance of a particular Pro residue based on the structure of a protein. Characterizing the role of specific Pro residues by calculating the isomerization energy in the framework of classical atomistic molecular modeling has already been attempted by Levitt [117]. However, this approach was based on energy minimization and energy evaluation of the isomer states and neglected exhaustive configurational sampling and entropic contributions. Assessing isomerization free energies rather

than energies requires exhaustive sampling and is hence computationally more demanding, but it could have important applications towards understanding the role of Pro residues [108]. For example, the magnitude of a Pro isomerization free energy may be related to the role of the Pro residue, i.e. connected to phenomena such as cis/trans heterogeneity, [118] molecular timing [119] or the importance of the isomer state in protein folding [120]. Hence, being able to accurately calculate Pro isomerization free energies for a wide range of proteins may have the potential to leverage our understanding of the structural and/or functional role of certain Pro residues. In general, structure-based in silico approaches may provide great insight for the explanation of protein function and specific biological mechanisms. Since the cis/trans isomerization is an inherently quantum-mechanical (QM) process, involving for example geometrical distortions in the transition state that transform the Pro nitrogen atom from a sp2-hybridized electronic state into a sp3-hybridized one, a QM description is in principle the most accurate way to investigate the complicated isomerization process [121]. However due to immense computational cost most of the QM studies are limited to the realm of small peptides with limited transferability to larger proteins. Adding even more complexity, Pro cis/trans isomerization has more than one relevant degree of freedom. It is not only the imide bond dihedral angle $\omega$ which is involved, but also the out-of-plane deformation of the imide nitrogen ("pyramidality"; Fig. 3.2 D) and the Pro backbone dihedral angle $\psi$, which all have a considerable effect on the isomerization barrier [122]. Considering the study of Pro cis/trans isomerization with classical atomistic molecular dynamics (MD) simulations, another complication is due to the fact that classical force-field descriptions exhibit large potential energies in the transition state along the cis/trans $\omega$ dihedral angle. This poses the risk of high forces which could drive the system into unfavorable regions of phase space when one tries to overcome this huge barrier height by means of certain advanced sampling methods such as Umbrella sampling (US) [123].

MD studies targeting Pro isomerization have to resort to advanced sampling methods due to the excessive timescales involved, which are beyond the reach of brute-force atomistic explicit-solvent simulations. In a study by Moradi et al. [124], the transition between and the free-energy landscapes of polyPro I (all-cis) and II (all-trans) helices in polyPro peptides was investigated using adaptively biased MD [125], enhanced by temperature [126, 127] and Hamiltonian replica exchange [89, 128] or nonequilibrium simulations in combination with the Crooks theorem [129]. Most simulations in this study were performed in implicit solvent. Another study [130] utilized accelerated MD [131] to sample cis/trans transitions of two adjacent Pro residues in an antibody, aiming to identify the preferred combi-

nation of isomer states. Gaussian accelerated MD [132] was employed in another study [133] to investigate cis/trans isomerization in various peptides, requiring a lowering of the peptide bond dihedral angle barrier to enhance transition rates. Free-energy profiles for Pro cis/trans isomerization in $\beta_2$-microglobulin were calculated in Ref. [134] using metadynamics [135] simulations, a technique that enhances sampling by gradually filling up the potential-energy surface in minima with additional bias functions (Fig. 3.2 A; top panel). However, a potential issue in calculating Pro cis/trans isomerization free energies with metadynamics is that, by sampling on a flattened potential energy surface, the focus on the dihedral angle values of interest (cis/trans states) is diminished over time. The Umbrella sampling (US) method, another commonly used technique, relies on bias potentials to concentrate configurational sampling in selected, potentially unfavorable areas of phase space (Fig. 3.2 A; bottom panel). However, in the context of Pro cis/trans isomerization, the application of strong harmonic potentials – necessary to overcome the transition barrier-along the $\omega$ dihedral angle and induce significant local structural changes can lead to the disruption of the native protein secondary structure. A third method, Hamiltonian replica exchange molecular dynamics (HREMD) [89, 126], involves running several replicas of the original system with different Hamiltonians simultaneously. These modifications aim to enhance sampling in higher replicas, facilitating the crossing of barriers. HREMD has been widely used to explore conformational transitions in peptides and proteins by introducing bias potentials on backbone ($\phi$, $\psi$) and side chain dihedral angles [136, 137].

In the present work, we introduce a novel advanced sampling method called $\omega$-bias potential replica exchange molecular dynamics ($\omega$BP-REMD) to overcome the challenges associated with the high energy barrier of isomerization. This method utilizes HREMD simulations with modifications to the Hamiltonian, including ($i$) decreased isomerization barrier heights in the replica ladder for enhanced sampling and ($ii$) a bespoke $\omega$-bias potential in all replicas for computational efficiency (Fig. 3.2 B). This approach addresses the issues related to energetically unfavorable transition states, allowing for an efficient and accurate determination of the cis/trans equilibrium of a given Pro residue. The method is based on the reweighted sampled probability distribution in the lower-most replica, where sampling is focused on physically relevant phase-space regions. Since the approach relies on a classical force-field description, simulation of large proteins is feasible. We compare isomerization free energies obtained from the $\omega$BP-REMD method with results from standard US. The systems studied involve tripeptides as well as the N2 domain of the gene-3-protein of phage fd [138].

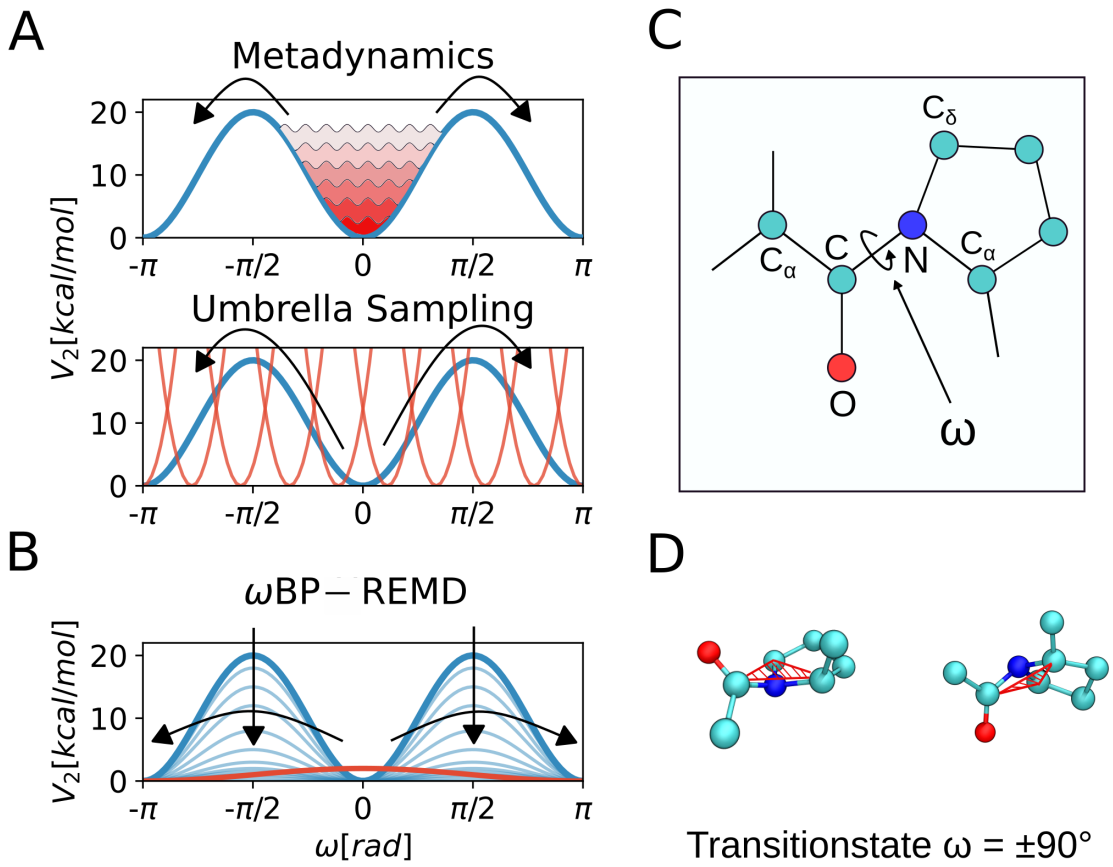



**Fig. 3.2.** Schematic depiction of advanced sampling schemes to enable the crossing of large potential-energy barriers. (A) Most advanced sampling techniques do not alter the potential-energy surface (PES) at the high-energy transition states. Metadynamics [135, 139] enhances sampling by gradually filling up the PES in minima with Gaussian functions. Umbrella sampling [123] applies harmonic bias potentials in order to force the system out of local minima. (B) In the HREMD method,[89, 140] several simulations (replicas) with modified Hamiltonians are employed. In the presented $\omega$BP-REMD method, the modification involves reduction of the barrier height of the $\omega$ dihedral angle energy profile (Eq. 3.1). This allows cis/trans transitions in higher replicas, which are propagated to lower replicas based on exchange of configurations according to the Metropolis criterion [90]. Furthermore, a small additional potential-energy term of multiplicity $n = 1$ for $\omega$ is introduced in all replicas (red line; Eq. 3.2). This is to achieve balanced sampling of cis/trans states especially in the lower replicas. (C) In the AMBER ff14SB forcefield[141] all four cosine terms along the $N_\chi = 4$ possible X-$C_{i-1}$-$N_i$-Y combinations contribute equally to the barrier height of the cis/trans isomerization energy profile of 20 kcal $\cdot$mol$^{-1}$ (Eq. 3.1). (D) Ball-and-stick representation of a Pro residue in the transition state of the isomerization. The plane formed by the three nitrogen-bound atoms is highlighted in red. Typical configurations for a positive (left) and negative (right) pyramidality (Eq. 3.21) are depicted. The out-of-plane movement of the nitrogen atom is an important degree of freedom during cis/trans isomerization.

## 3.2 Material and Methods

For all MD simulations, the AMBER20 software suite and the AMBER ff14SB force field [141, 142] were used. Water was treated explicitly with the TIP3P water model[143, 144].

### 3.2.1 Backbone Dihedral Angle Potential in the AMBER Force-Field Series

In the AMBER ff14SB force field, as well as in precursory forcefields, the energy profile $E_{Dih}(\Phi;\Phi_{\chi[c]})$ for the torsion $\Phi$ around a backbone peptide bond is given by a sum of Fourier series [141],

$$E_{Dih}(\Phi;\Phi_{\chi[c]}) = \sum_{c=1}^{N_\chi}\sum_{n=1}^{2} V_{\chi[c],n}[1+\cos(n\Phi_{\chi[c]}+\gamma_{\chi[c],n})] \tag{3.1}$$

where $V_{\chi[c],n}$, $\gamma_{\chi[c],n}$, $\Phi_{\chi[c]}$, and $n$ denote force constants, phases, dihedral angle values and multiplicity, respectively, of the dihedral angle $\chi[c]$.
The first sum in Eq. 3.1 runs over the $N_\chi$ dihedral angles $\chi[c]$ formed by the $N_\chi$ four-atom sets given by the first neighbors of nitrogen and carbon atoms involved in the peptide bond (Fig. 3.1). Modeling a dihedral angle potential by an in principle redundant set of all possible $N_\chi$ dihedral angle definitions around the concerned bond rather than only one thereof is a convention used in the AMBER series of forcefields. Choosing only one of the possible dihedral angle definitions along with a corresponding multiple of the force constant, as done in certain other forcefields, may be more computationally efficient. However, the advantage of the redundant specification is that no arbitrary choices which may lead to asymmetries in the system have to be made.
The second sum in Eq. 3.1 is a Fourier series and thus contributes with terms of different multiplicities $n$. The force constants and phase parameters depend on the involved atom sets (dihedral angles $\chi[c]$) and the multiplicity, which is indicated by corresponding subscripts. In general, the adjustment of the energy profile via $N_\chi$ Fourier series allows to describe complex variations of the potential energy with bond rotation, e.g. due to orbital or steric effects, and to account for the particular treatment of 1-4 nonbonded interactions. This is why the corresponding torsion profile parameters are typically fitted last when parameterizing force fields. It is a

key assumption that individual dihedral angle energy profiles are uncoupled, i.e. do not explicitly depend on neighboring dihedral angles [141].
Note that the first two terms ($n = 1, 2$) in the Fourier series (Eq. 3.1) are sufficient to represent the peptide bond dihedral angle energy profile. The onefold rotational term ($n = 1$, $\gamma_{\chi[c],1} = 0$) with a maximum at a dihedral angle of 0 determines the energy difference between the trans and cis isomers, whereas the second term ($n = 2$, $\gamma_{\chi[c],2} = \pi$) reproduces the twofold symmetry of the peptide bond dihedral angle energy profile with minima at dihedral angles of $\pm\pi$ and 0. In particular, the force constant pertaining to the twofold term, $V_{\chi[c],2}$, essentially determines the barriers to rotation about the C-N peptide bond and the force constant pertaining to the onefold term, $V_{\chi[c],1}$, is responsible for the correct cis/trans equilibrium.
In the AMBER force field series, the parameters $V_{\chi[c],1}$ and $V_{\chi[c],2}$ are specified for each set of four atoms (i.e., X-$C_{i-1}$-$N_i$-Y where X can be a $C_\alpha$ carbon atom or a carbonyl oxygen atom and Y can be a $C_\alpha$ carbon atom or an amide hydrogen atom) that define the $\omega$-dihedral angle (Fig. 3.1) and subscripts $i-1$, $i$ denote the amino acid index. For non-Pro peptide bonds terms associated with both $V_{\chi[c],1}$ and $V_{\chi[c],2}$ are required to accurately model the dihedral angle energy profile (Eq. 3.1) [145]. However, in case of Pro containing peptide bonds (Xaa-Pro) only the second term $V_{\chi[c],2}$ is necessary because the comparatively small energy differences between cis and trans isomers can be reproduced with only the twofold rotational term [145]. That is, $V_{\chi[c],1} = 0$ in this case (Table 3.1). Each of the four possible combinations X-$C_{i-1}$-$N_i$-Y (see above and Fig. 3.2 B) contributes with a value of $V_{\chi[c],2} = 2.5$ kcal·mol$^{-1}$ in the AMBER ff14SB forcefield[141] (Eq. 3.1; Table 3.1). Taken together, this amounts to an effective torsional energy barrier of 20 kcal·mol$^{-1}$. Note, in this context, that due to the planarity of the nitrogen atom and its bound atoms as well as of the carbonyl carbon atom and its bound atoms, pairs among the four dihedral angles are either equal or differ by $\pi$. A difference of $\pi$ does, however, not lead to a difference in the $V_{\chi[c],2}$-associated potential-energy contribution in Eq. 3.1, because the twofold cosine terms corresponding to $V_{\chi[c],2}$ constructively add up for torsions differing by $\pi$. Note also that the value of $V_{\chi[c],2} = 2.5$ kcal·mol$^{-1}$ has been seen to render the free-energy barrier between the cis and trans states too low compared to experimental data [145]. Nevertheless, we chose to employ this value in the lowermost replica because the focus of the present study is the calculation of isomerization free energies rather than isomerization rates. In the following, we refer to the standard definition of $\omega$, meaning the $C_{\alpha,i-1}$-$C_{i-1}$-$N_i$-$C_{\alpha,i}$ dihedral angle, but note that the forcefield accounts for the corresponding isomerization barrier height of 20 kcal·mol$^{-1}$ by a sum involving the four possible X-$C_{i-1}$-$N_i$-Y dihedral angles.

**Tab. 3.1.** $\omega$ dihedral angle forcefield parameters $V_{\chi[c],n}$ and $\gamma_{\chi[c],n}$ (Eq. 3.1) for a Pro-containing peptide bond between residues $i-1$ and $i$, where the latter is the Pro residue, in the AMBER ff14SB forcefield.[141] The four contributing dihedral angles $\chi[c]$ are listed and the first two or last two atoms of the dihedral angle pertain to residue $i-1$ or $i$, respectively.

| $\chi[c]$ | $V_{\chi[c],1}$ [kcal·mol$^{-1}$] | $\gamma_{\chi[c],1}$ | $V_{\chi[c],2}$ [kcal·mol$^{-1}$] | $\gamma_{\chi[c],2}$ |
|---|---|---|---|---|
| O-C-N-$C_\delta$ | 0 | 0 | 2.5 | $\pi$ |
| O-C-N-$C_\alpha$ | 0 | 0 | 2.5 | $\pi$ |
| $C_\alpha$-C-N-$C_\delta$ | 0 | 0 | 2.5 | $\pi$ |
| $C_\alpha$-C-N-$C_\alpha$ | 0 | 0 | 2.5 | $\pi$ |

### 3.2.2 Modified Hamiltonian in the $\omega$BP-REMD Method

For the $\omega$BP-REMD simulations, in this case, 12 replicas differing in the barrier height ($V_{\chi[c],2}$) of the $\omega$ dihedral angle potential (Eq. 3.1) were created for each simulated system. Note that, although referring to the $\omega$ dihedral angle, this actually involves the four dihedral angles X-$C_{i-1}$-$N_i$-Y surrounding the peptide bond (Fig. 3.2 C). Equidistant spacing from $V_{\chi[c],2} = 0$ up to $V_{\chi[c],2} = 2.5$ kcal·mol$^{-1}$ was chosen for each of the four torsion terms, resulting in the effective potentials shown in Fig. 3.2 B. The replica with $V_{\chi[c],2} = 2.5$ kcal·mol$^{-1}$, is referred to as replica 1 or the lowest replica in the following. It corresponds to the $V_{\chi[c],2}$-value used in the original AMBER ff14SB force field[141] (Table 3.1). However, it differs from the original force-field dihedral angle potential in that a potential-energy term $W_1(\omega)$ with multiplicity $n = 1$ is added along the $C_{\alpha,i-1}$-$C_{i-1}$-$N_i$-$C_{\alpha,i}$ dihedral angle (Fig 3.2 B; red line),

$$W_1(\omega) = V_{\omega,1}[1 + \cos(\omega + \gamma_{\omega,1})] \quad . \tag{3.2}$$

The inclusion of this term allows adjustment of the cis/trans equilibrium and can be used to steer configurational sampling out of the trans state when a phase $\gamma_{\omega,1} = \pi$ is used. To destabilize the cis state, accordingly, $\gamma_{\omega,1} = 0$. We used a value of 1.0 kcal·mol$^{-1}$ for the barrier height $V_{\omega,1}$ (Eq. 3.2) in all simulations in order to achieve approximately balanced cis/trans sampling. This choice was motivated by analysis of the isomerization free energy based on a preliminary free-energy profile in absence of the $W_1(\omega)$ bias potential.
In summary, the dihedral angle potential for the $\omega$ dihedral angle in the $\omega$BP-REMD simulations for replica 1 is composed of the original AMBER ff14SB force-field

terms[141] corresponding to Eq. 3.1 and the additional small bias potential in Eq. 3.2,

$$E_{Dih}(\omega;\Phi_{\chi[c]}) = \sum_{c=1}^{4} V_{\chi[c],2}[1+\cos(2\Phi_{\chi[c]}+\gamma_{\chi[c],2})] + V_{\omega,1}[1+\cos(\omega+\gamma_{\omega,1})] \quad , \text{(3.3)}$$

where $V_{\chi[c],2} = 2.5$ kcal·mol$^{-1}$, $\gamma_{\chi[c],2} = \pi$ and $V_{\omega,1} = 1.0$ kcal·mol$^{-1}$. For the remaining 11 replicas $r$,

$$E_{Dih}(\omega;\Phi_{\chi[c]}) = \sum_{c=1}^{4} V_{\chi[c],2;r}[1+\cos(2\Phi_{\chi[c]}+\gamma_{\chi[c],2})] + V_{\omega,1}[1+\cos(\omega+\gamma_{\omega,1})] \quad , \text{(3.4)}$$

where $V_{\chi[c],2;r} = 0$, 0.227, ..., 2.273 kcal·mol$^{-1}$ in the order of decreasing replica index $r = 12, 11, ..., 2$. The phase $\gamma_{\omega,1}$ in the added onefold term is either set to 0 or $\pi$ if destabilization of the cis or trans state, respectively, is desired.

During the MD production run, configurational exchange attempts between the replicas were conducted every 250 steps based on the Metropolis criterion [90]. This means that the exchange probability $w_{i,j}$ between neighboring replicas $i$ and $j$ is given by

$$w_{i,j} = \min(1, \exp^{-\beta\Delta E_{i,j}}) \quad , \text{(3.5)}$$

where $\beta = (k_B T)^{-1}$, where $k_B$ is the Boltzmann constant and $T$ the absolute temperature, $\Delta E_{i,j}$ is the potential energy difference

$$\Delta E_{i,j} = E_j(\mathbf{r}^N;i) - E_i(\mathbf{r}^N;i) + E_i(\mathbf{r}^N;j) - E_j(\mathbf{r}^N;j) \quad , \text{(3.6)}$$

$\mathbf{r}^N$ is the Cartesian coordinate vector containing the coordinates of all $N$ atoms in the system and $E_i(\mathbf{r}^N;j)$ is the potential energy calculated according to the Hamiltonian of replica $i$ using the coordinates $\mathbf{r}^N$ of a configuration that was obtained during the simulation of replica $j$. Due to the Metropolis criterion in Eq. 3.5, a fraction of attempted configurational exchanges will actually not be carried out. This ensures canonical sampling of configurational ensembles in all replicas. In the present work, the difference in the Hamiltonians between replicas concerns only the applied bias potentials, i.e. Eq. 3.3 (for replica 1) and Eq. 3.4 (for the remaining eleven replicas). With the values reported above for $V_{\chi[c],2}$, $V_{\omega,1}$

and $V_{\chi[c],2;r}$, the difference in the Hamiltonian between neighboring replicas was sufficiently small to allow considerable overlap of sampled configurations in terms of energy (Supporting Information (SI) Fig. A.1 and A.2) and thus frequent configurational exchange. For example, the mean exchange acceptance rate was 79.8% for peptide DPV and of similarly high magnitude for the other simulated systems (SI Tables A.1 and A.2). Inspection of the actual exchanges between replicas showed that configurational exchanges occur throughout all successive replicas and hence configurations can indeed propagate downwards throughout the entire replica ladder (SI Fig. A.3 and A.4), which entails the desired effect of enhancing the sampling of $\omega$-values in replica 1. The free-energy profile along $\omega$, i.e. the potential of mean force (PMF)[146], and hence the isomerization free energy, was obtained from the sampled density of states after reweighting. This is detailed in Section 3.2.5.

### 3.2.3 Umbrella Sampling Simulations

US simulations were performed as an alternative method to verify the reliability of the $\omega$BP-REMD method. The $C_{\alpha,i-1}$-$C_{i-1}$-$N_i$-$C_{\alpha,i}$ dihedral angle served as reaction coordinate $\omega$. Sets of 24 US windows $i$ biased by harmonic potentials

$$E_{bias,i} = k(\omega - \omega_{0,i})^2 \tag{3.7}$$

with force constants $k = 30$ kcal·mol$^{-1}$· rad$^{-2}$ were generated with the minima of the harmonic potentials located at $\omega_{0,i} = -\pi...\pi$ in equidistant spacings of $0.086\pi$ (Fig. 3.2 A; bottom panel) . If required, additional windows with a force constant of 50 kcal·mol$^{-1}$· rad$^{-2}$ were inserted to improve the overlap between neighboring windows. Note that the reaction coordinate refers to only one particular description of the $\omega$ dihedral angle, which is in contrast to the potentials used in the $\omega$BP-REMD simulations acting on the four dihedral angle descriptions (Eq 3.3, 3.4, with $N_\chi = 4$). The free-energy profile along $\omega$ and hence the isomerization free energy can be obtained from the sampled density of states after reweighting. It was calculated as described in Section 3.2.5.

### 3.2.4 Simulation Details

The CUDA accelerated PMEMD version[147] of the AMBER20 software package[148] was used for all simulations. The applicability of the $\omega$BP-REMD method was tested on a variety of Pro containing tripeptides with N-terminal acetyl and C-terminal N-methyl amide capping groups. Here, all structures were built using the AmberTools LEaP module [148]. The method was further applied to the isomerization of a Pro residue in a folded protein environment, the N2 domain of the gene-3-protein (G3P) of phage fd and certain mutants thereof. The corresponding simulations were based on the PDB structure 1G3P [138]. The N2 domain is a small, 104-residue protein. It features a biologically-important $\beta$-hairpin loop where the Pro of interest, here referred to as Pro161 to be consistent with previous experimental studies [120], resides in the solvent-exposed loop region, at the connection of two $\beta$ strands. Protonation states of titrable residues were predicted via Poisson-Boltzmann calculations, using the free open-source webserver H++ [149]. Each of the peptides or proteins was solvated using explicit TIP3P water molecules in periodic octahedral boxes [143, 144]. Minimum distances of 12 Å to the box boundaries were chosen for the larger protein system and 18 Å for the tripeptide systems, respectively. Sodium and chloride ions were added to neutralize the systems and reach an ion concentration of 100 mM. After an energy minimization of 2000 steps of steepest descent, the tripeptide systems were heated up to 300 K during three simulations of in total 120 ps length, while keeping positional restraints with a force constant of 25 kcal·mol$^{-1}$ · Å$^{-2}$ on the solute non-hydrogen atoms. To maintain the respective desired temperatures, the Berendsen weak-coupling algorithm [84] was used with a coupling time of 0.1 ps. All systems concerning the N2 domain of G3P were heated to up to 288 K in the same way in order to reproduce the experimental conditions[120]. In five subsequent simulations of in total 1000 ps length, the positional restraints were gradually released and the density was equilibrated at a pressure of 1 bar, using the Berendsen weak-coupling algorithm [84] with a coupling time of 2.5 ps. All bonds involving hydrogen atoms were kept at the optimal bond length using the SHAKE algorithm [82]. Hydrogen mass repartitioning was used allowing a time step of 4 fs [83]. All production simulations were performed at constant temperature and constant volume at the equilibrated density. The temperature was held constant with the Berendsen weak-coupling algorithm [84] with a coupling time of 2.5 ps. A non-bonded cutoff of 9 Å was used in all simulations to truncate Lennard-Jones interactions and real-space interactions in the PME method [77]. Atom coordinates were written out every 40 ps.

### 3.2.5 Analysis of Simulations

For the $\omega$BP-REMD simulations a reweighting scheme according to the Zwanzig equation [150] was used to estimate the cis/trans isomerization free energy $\Delta G$. This was done in two possible ways. In a first way (Section 3.2.5), the cis/trans equilibrium was estimated based on sampling in replica 1 only, where the barrier height of the cis/trans isomerization energy profile of 20 kcal·mol$^{-1}$ corresponds to the original AMBER ff14SB force-field description[141]. Reweighting[150] of configurations is required to account for the effect of the introduced small bias potential on the cis or trans state (Eq. 3.2). In a second way (Section 3.2.5), the cis/trans equilibrium was estimated based on sampling in all replicas. A Weighted Histogram Analysis (WHAM) approach[103] was used to unbias and combine the corresponding sampled densities of states. It was implemented in a python script [@151].

#### Cis/trans Isomerization Free Energy from Replica 1

As described in Section 3.2.2, replica 1 represents Boltzmann sampling according to the original AMBER ff14SB force field[141] supplemented by the additional bias potential of multiplicity $n = 1$ acting on the $\omega$ dihedral angle (Eq. 3.2). Thus, the sampling corresponds to the biased probability distribution [34]

$$\tilde{P}_1(\omega) = \exp^{\beta G_1} \int d\mathbf{r}^N \exp^{-\beta[U(\mathbf{r}^N)+W_1(\omega(\mathbf{r}^N))]} \delta(\omega(\mathbf{r}^N) - \omega) \quad , \tag{3.8}$$

where $G_1$ is, to within an additive constant, the free energy associated with the biased sampling in replica 1, i.e.

$$G_1 = -k_B T \ln \left[ \int d\mathbf{r}^N \exp^{-\beta[U(\mathbf{r}^N)+W_1(\omega(\mathbf{r}^N))]} \right] \quad , \tag{3.9}$$

where $\omega(\mathbf{r}^N)$ denotes the value of $\omega$ in a particular configuration $\mathbf{r}^N$ and where $\delta(x)$ is the Dirac delta function. $W_1(\omega(\mathbf{r}^N))$ denotes the introduced bias (Eq. 3.2) and $U(\mathbf{r}^N)$ is the physical potential energy of configuration $\mathbf{r}^N$. Note that kinetic energy contributions are omitted throughout because here only a change in a configurational variable, $\omega(\mathbf{r}^N)$, is considered. Note also that pressure-volume contributions are neglected due to their generally small magnitude. The probability $\tilde{P}_1(\omega)$ is normalized, i.e. $G_1$ (Eq. 3.9) is in practice absorbed by a normalization

constant. $\tilde{P}_1(\omega)$ was extracted from the simulations via Kernel Density Estimation (KDE) with a von-Mises kernel,

$$f(x;\mu,\kappa) = \frac{\exp\left(\kappa\cos\left(x-\mu\right)\right)}{2\pi I_0(\kappa)} \tag{3.10}$$

where $\mu$ denotes the location of the maximum probability, $\kappa^{-1}$ corresponds to the variance or bandwidth $h$ of the kernel, $I_0$ is the modified Bessel function of the first kind of order 0 and $I_0(\kappa)$ fulfils the role of a normalization constant. The optimal bandwidth was selected by Maximum Likelihood Cross-Validation (MLCV). In this approach, the logarithm of the likelihood $\log\prod_{i=1}^{n} f_h(X_i)$ is maximized, where $X_i$ are the sampled $\omega$ dihedral angles and $f_h$ denotes the von-Mises distribution (Eq. 3.10) for a given value of $h = \kappa^{-1}$. Here, $K$-fold cross validation ($K = 100$) was employed to avoid overfitting (SI Fig. A.5 and A.6). Note that kernel smoothing for $\tilde{P}_1(\omega)$ does not necessarily have to be done using the von-Mises distribution but could for example also rely on Gaussian kernels [152]. The unbiased (reweighted) probability distribution $P_1(\omega)$ was evaluated based on the biased probability distribution $\tilde{P}_1(\omega)$ (Eq. 3.8) via [34]

$$P_1(\omega) = \exp^{-\beta(G_1-G_0)}\exp^{\beta W_1(\omega)}\tilde{P}_1(\omega) \quad , \tag{3.11}$$

where $G_0$ relates to the free energy associated with the sampling in the unbiased system,

$$G_0 = -k_B T\ln\left[\int d\mathbf{r}^N \exp^{-\beta U(\mathbf{r}^N)}\right] \quad . \tag{3.12}$$

It follows from Eqs. 3.9, 4.3 and 4.5 that

$$P_1(\omega) = \tilde{P}_1(\omega)\exp^{\beta W_1(\omega)}\langle\exp^{-\beta W_1(\mathbf{r})}\rangle_{unbiased} \quad , \tag{3.13}$$

where the ensemble average $\langle\ldots\rangle_{unbiased}$ is taken over an unbiased trajectory, or, equivalently,

$$P_1(\omega) = \tilde{P}_1(\omega)\exp^{\beta W_1(\omega)}\langle\exp^{\beta W_1(\mathbf{r})}\rangle^{-1}_{biased} \quad , \tag{3.14}$$

where the ensemble average is taken over the biased trajectory. The free-energy profile along $\omega$ is directly linked to this probability density by

$$G(\omega) = -k_B T\ln(P_1(\omega)) \tag{3.15}$$

and the free energy of isomerization $\Delta G$ was calculated by subtraction of the free energies corresponding to the cis and trans states, $G_{cis}$ and $G_{trans}$, respectively,

$$\Delta G = G_{cis} - G_{trans} \quad , \tag{3.16}$$

where $G_{cis}$ and $G_{trans}$ were taken as the free-energy values in the minima of the free-energy profile. Statistical errors on $\Delta G$ were evaluated as the standard deviation of free energies obtained from dividing the total simulation length into 50 pieces of successively increasing length by 2 ns.

### Cis/Trans Isomerization Free Energy Based on all Replicas

The configurational sampling in replica 1 is $(i)$ focused on the cis and trans states of the considered peptide bond and $(ii)$ shared approximately equally between the two states (Fig. 3.3 and 3.6). Thus, the presented $\omega$BP-REMD method (Section 3.2.2) and the analysis based on replica 1 (Section 3.2.5) are a computationally efficient strategy to calculate cis/trans isomerization free energies accurately and with high statistical precision. Since the original barrier height of the $\omega$ dihedral angle energy profile is unaltered in replica 1 ($V_{\chi[c],2} = 2.5$ kcal·mol$^{-1}$ ; Eq. 3.1 and Table 3.1) and very high (20 kcal·mol$^{-1}$; Eq.3.1), the transition state regime of the isomerization has very low Boltzmann factors and is essentially unsampled in replica 1. Therefore, no information about the free-energy profile near the transition state is obtained based on replica 1. However, higher replicas permit sampling of the transition state regime, i.e. $\omega$ dihedral angle values of about $\pm\pi/2$. For reasons of completeness, the analysis was extended by taking into account the sampling of all replicas and calculating the PMF with the WHAM method[103], meaning that the biased replicas 2-12 ($V_{\chi[c],2} < 2.5$ kcal·mol$^{-1}$; Section 3.2.2) are also taken into account. The biased probability distribution in replica $k$ is [34]

$$\tilde{P}_k(\omega) = \exp^{\beta G_k} \int d\mathbf{r}^N \exp^{-\beta[U(\mathbf{r}^N)+W_k(\omega(\mathbf{r}^N))]} \delta(\omega(\mathbf{r}^N) - \omega) \quad , \tag{3.17}$$

where $W_k(\omega(\mathbf{r}^N))$ is the bias potential in replica $k$, i.e. the difference between the modified dihedral angle energy profile in replica $k$ (Eq. 3.4) with $V_{\chi[c],2;r} = 2.273, ...0.227, 0$ kcal·mol$^{-1}$ in replicas $r = 2, 3, ...11, 12$, respectively, and the origi-

nal one (Eq. 3.1). The free energy associated with the biased sampling in replica $k$ is, to within an additive constant,

$$G_k = -k_B T \ln \left[ \int d\mathbf{r}^N \exp^{-\beta[U(\mathbf{r}^N)+W_k(\omega(\mathbf{r}^N))]} \right] \quad , \tag{3.18}$$

and is in practice absorbed by a normalization constant. Corresponding unbiased (reweighted) probabilities are

$$P_k(\omega) = \tilde{P}_k(\omega) \exp^{\beta W_k(\omega)} \langle \exp^{\beta W_k(\omega(\mathbf{r}^N))} \rangle^{-1}_{biased} \tag{3.19}$$

and according to the WHAM method[103], the full probability $P(\omega)$ along $\omega$ is determined as a linear combination of $P_k(\omega)$ with coefficients determined to minimize the variance in $P(\omega)$. The resulting free-energy profile along $\omega$ is then

$$G(\omega) = -k_B T \ln(P(\omega)) \quad . \tag{3.20}$$

Again, the isomerization free energy was calculated according to Eq. 3.16. Note that for the $\omega$BP-REMD simulations the reweighting procedure is slightly artificial for the bins in the transition state regime because the four X-$C_{i-1}$- $N_i$-Y dihedral angles (Fig. 3.2 C) are not perfectly correlated in these regions and in the present work, the $\omega$ dihedral angle, rather than all four actual dihedral angles along the peptide bond, is used to calculate the bias potentials $W_k(\omega(\mathbf{r}^N))$ (Eq.3.18). The corresponding reweighting procedure and resulting unbiased (reweighted) probabilities (Eq. 3.19) are an estimate based on the assumption that all four dihedral angles are equal to $\omega(\mathbf{r}^N)$ in all sampled configurations $\mathbf{r}^N$. This assumption is, however, not justified when significant deviations from planarity occur in the carbonyl group and the around the nitrogen atom, because then the differences between the four dihedral angles X-$C_{i-1}$- $N_i$-Y no more evaluate to exactly 0 or $\pi$. Especially in the transition state region, deviations from planarity in these groups can occur (Section 6.2.4). However, an accurate depiction of the transition states in the isomerization process is not feasible through a classical force field description, nor was it the objective of the current study. This aspect will be revisited later.

### Cis/Trans Isomerization Free Energy Based on US

For analysis of the US simulations, PMFs along the $\omega$ dihedral angle were calculated with the WHAM method [103]. The isomerization free energy (Eq. 3.16) was

calculated based on the free-energy values in the minima of the free-energy profile. Statistical errors on $\Delta G$ were evaluated as the standard deviation of free energies obtained from dividing the total simulation length into 50 pieces of successively increasing length by 2 ns.

#### Structural Analyses

One way to quantify the misalignment between the $C_{\alpha,i-1}$ -$C_{i-1}$-$N_i$-$C_{\alpha,i}$ and the $C_{\alpha,i-1}$ -$C_{i-1}$-$N_i$-$C_{\delta,i}$ dihedral angles is the out-of-plane deformation of the imide nitrogen atom. This so-called pyramidality $p$ is known to be an important degree of freedom in the transition state and is defined as

$$p = \det |\mathbf{e1e2e3}| \tag{3.21}$$

where **e1**, **e2** and **e3** are unit vectors with directions pointing from N to $C_\delta$, $C_\alpha$ and C, respectively (Fig. 3.2 C, D). Analysis of hydrogen bonding was done by defining a hydrogen bond between an acceptor heavy atom A and a hydrogen atom H attached to a donor heavy atom D when the distance between atoms A and D was below a threshold of 3 Å and the A-H-D angle was above a threshold of 135°. Processing of the generated trajectories was done using Cpptraj and Pytraj [153]. Visualization and depiction of molecular conformations was done with VMD [154].

## 3.3 Results and Discussion

### 3.3.1 Tripeptide Systems

The performance of the $\omega$BP-REMD method was evaluated for nine Pro-containing tripeptides (DPV, NPA, NPL, NPN, NPS, PPG, QPL, TPV and YPP). The results were compared to US along the $\omega$ dihedral angle, US being considered the standard method to obtain the free-energy profile along a given reaction coordinate. The tripeptide systems are comparably small and were hence expected to show limited sampling issues and allow robust comparison between the two methods. Furthermore, the cis/trans equilibria of the tripeptides are of biological interest, as they serve as a nearest-neighbor model of the unfolded state of larger proteins.

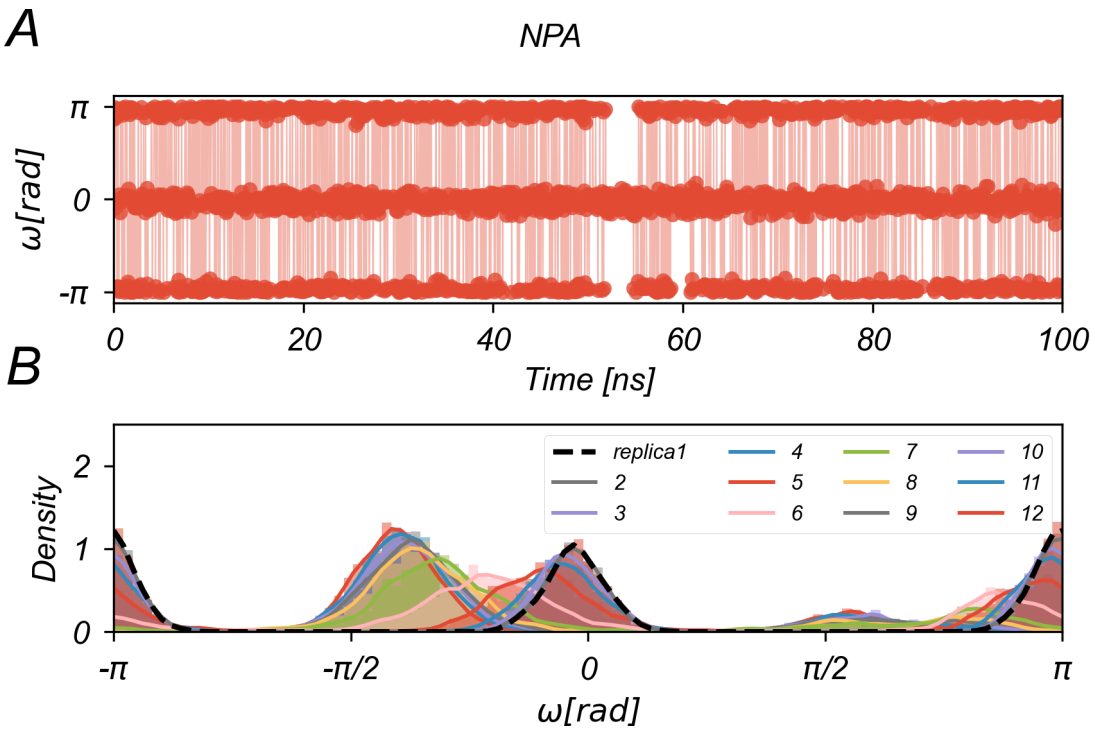



**Fig. 3.3.** Cis/trans isomerization in peptide NPA investigated with the $\omega$BP-REMD method (Section 3.2.2). Data for the other tripeptides is reported in SI Fig. A.7 - A.15. (A) Sampling of the $\omega$ dihedral angle in replica 1, where the barrier height of the dihedral angle energy profile amounts to 20 kcal·mol$^{-1}$ (Eq. 3.3), during a simulation of 100 ns length. (B) Sampling of the $\omega$ dihedral angle in all 12 replicas depicted by the probability density along $\omega$. The probability density for replica 1, presenting balanced cis/trans sampling, is highlighted with the black dashed line. Transitions between the cis ($\omega = 0$) and trans ($\omega = \pm\pi$) states over the barrier at $\omega = -\pi/2$ are preferred for all tripeptide systems.

**Tab. 3.2.** Cis/trans (including trans/cis) exchange rates of the central Pro residue for the considered tripeptides and of Pro161 for the N2 domain of protein G3P (wildtype and protein mutants) in replica 1 of the $\omega$BP-REMD method. The simulation length was 100 ns (tripeptides) or 400 ns (proteins). Frames were written to file every 40 ps.

| Tripeptide, cis/trans exchanges of Pro2 [ns$^{-1}$] | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| DPV | NPA | NPN | NPS | PPG | QPL | TPV | YPP | NPL |
| 10.90 | 10.58 | 9.96 | 10.92 | 10.30 | 9.10 | 10.50 | 4.48 | 9.96 |

| Protein, cis/trans exchanges of Pro161 [ns$^{-1}$] | | | | |
|---|---|---|---|---|
| Wildtype | G158A | K163A | L125A | Y165A |
| 9.98 | 4.75 | 9.03 | 3.29 | 5.25 |

For all investigated tripeptide systems, the $\omega$BP-REMD method was capable of introducing frequent cis/trans transitions in replica 1, where the original AMBER ff14SB dihedral angle isomerization barrier height[141] of 20 kcal·mol$^{-1}$ is effective (tripeptide NPA in Fig. 3.3 A; data for other tripeptides in SI Fig. A.7 - A.15 A). Both states are frequently visited in the course of 100 ns MD simulation for all systems with excellent exchange rates in the range of 4.5 - 10.9 ns$^{-1}$ (Table 3.2).

Thus, the $\omega$BP-REMD method has the capability of sampling both cis and trans states even in presence of the full dihedral angle energy barrier (replica 1; Eq. 3.3). This is done by propagating energetically favorable configurations, which have surmounted the transition states (at $\pm\pi/2$) in higher replicas. Here, it shows that cis/trans transitions for the tripeptides are predominantly accomplished via the path $0 \rightarrow -\pi$. This finding is in agreement with QM calculations performed by Yonezawa et al.[121] where it was shown that the hydration of the Pro nitrogen atom stabilizes the negative pyramidal conformation.

The amount of time spent in the cis and trans states of the $\omega$ dihedral angle of the considered tripeptides is approximately equal. This is due to the additional onefold bias potential ($W_1(\omega)$; Eq. 3.2) where $V_{\omega,1} = 1$ kcal·mol$^{-1}$ was sufficient to drive the system out of the otherwise favorable trans state and achieve approximately equal cis/trans sampling in replica 1 (tripeptide NPA in Fig. 3.3 B; data for other tripeptides in SI Fig. A.7 - A.15 B). Configurational sampling that is focused on the important regions of phase space (here, the cis/trans states) along with approximately equal sampling in the free energy minima of interest, is a desired feature of advanced sampling methods because it enhances the statistical efficiency of the reweighting. A mentionable exception is tripeptide YPP, where the equilibrium is shifted more towards the cis state (approximately 88% sampling in the cis state during 100 ns in the lowest replica; SI Fig. A.15). The comparably

favored sampling of the cis state is also reflected in a slightly reduced cis/trans exchange rate of 4.5 $ns^{-1}$ compared to the other tripeptides (Table 3.2). Due to the rather low free-energy difference between the cis and trans states of tripeptide YPP (see below and Fig. 3.4 C), a lower magnitude of $V_{\omega,1}$, i.e. less penalty on the trans state, may allow less sampling of the cis state in this special case.

In general, we observe that replicas with reduced barrier heights (i.e. replicas 2-12) appear to have a clear tendency to shift the cis state more towards negative $\omega$ dihedral angle values (with peaks mostly around -0.4 $\pi$, Fig. 3.3 B). Again, only YPP deviates with more frequent sampling in regions near $\pi/2$ (SI Fig. A.15).

The free energies of cis/trans isomerization were computed from probability distributions sampled along the $\omega$ dihedral angle. This involved extracting the free energy values using Equation 3.16.

For the $\omega$BP-REMD simulations, this is done by reweighting the sampled configurations in replica 1 according to Eq. 4.3, which accounts for the added bias potential $W_1(\omega)$ (Eq. 3.2). The resulting free-energy profile is incomplete along $\omega$ because due to the unaltered magnitude of $V_{\chi[c],2}$ in replica 1, virtually no sampling is done in the transition state regime on the considered simulation time scale (Fig. 3.4 A; green line for peptide NPA; data for other tripeptides in SI Fig. A.16). This leads to a spurious overestimation of the free energy in the transition-state regions (Fig. 3.4 A; green line tending to $+\infty$ for $\pm\pi/2$). Nevertheless, the relevant phase-space regions, namely the cis and trans states, are adequately sampled with favorable Boltzmann weights, allowing for meaningful calculations of the corresponding free energies. For all tripeptides, the obtained isomerization free energies agree very well with those from US, with a maximum deviation of 0.28 $kcal{\cdot}mol^{-1}$ for peptide YPP (Fig. 3.4 C). Note that peptide YPP presents the lowest isomerization free energy (0.96 $kcal{\cdot}mol^{-1}$ from $\omega$BP-REMD; Fig. 3.4 C). Low isomerization free energies are often observed in the context of cis/trans heterogeneity [118], a phenomenon, where both cis and trans states of a Pro residue can occur in a native protein fold. In fact, the YPP-motif was found before to show cis/trans heterogeneity in a certain antibody loop [155].
The dependence of calculated cis/trans equilibria on simulation length was investigated by considering subsets of the sampled trajectories. Running isomerization free energy estimates from both the $\omega$BP-REMD and US method converge fairly well after roughly 25 ns (Fig. 3.4 D, SI Fig. A.17).

The analysis of the $\omega$BP-REMD simulations can be extended by taking into account the sampling of all 12 replicas (Section 3.2.5). Since a larger range of $\omega$ dihedral

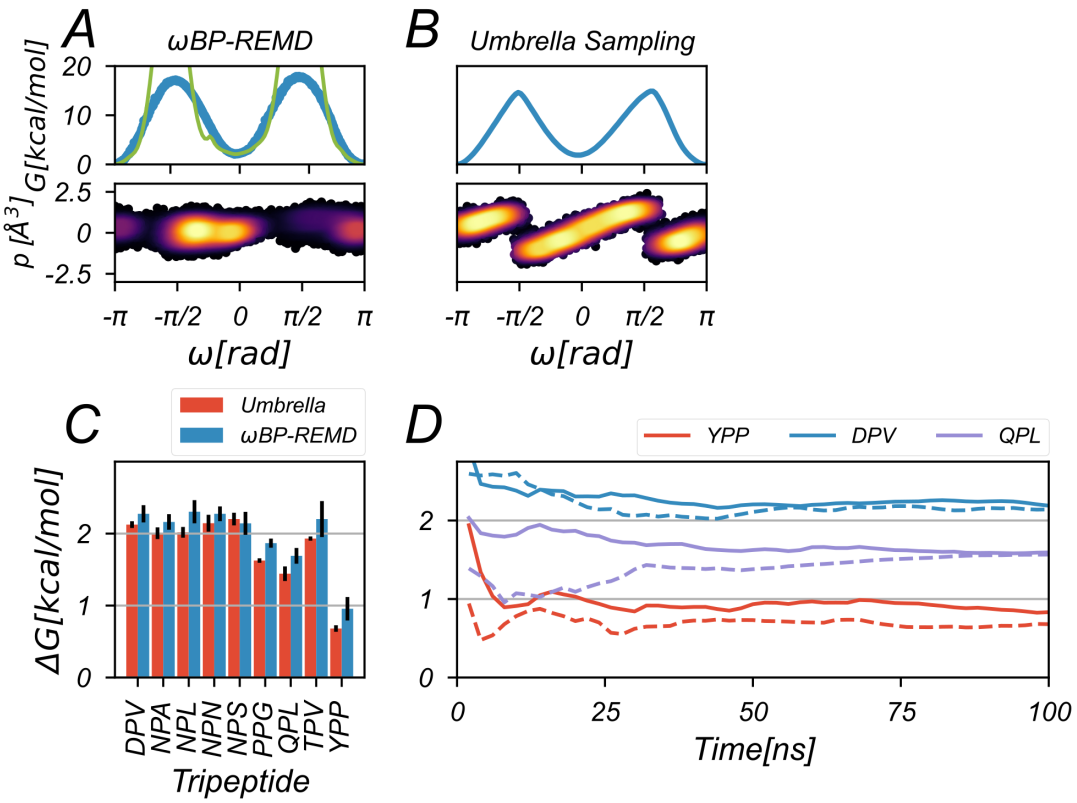



**Fig. 3.4.** Free energy of cis/trans isomerization. (A) Free-energy profile $G$ for peptide NPA based on the $\omega$BP-REMD method (top) and probability density of the pyramidality $p$ (bottom; Eq. 3.21) along the $\omega$ dihedral angle. The latter is not reweighted and includes data from all replicas. The green line indicates free energies calculated from unbiased (reweighted) probabilities based on replica 1 (Eq. 4.3; Section 3.2.5), where the transition state regime is essentially unsampled. The blue dots indicate the free-energy profile calculated from unbiased (reweighted) probabilities based on all replicas (Eq. 3.19; Section 3.2.5) and the WHAM method[103]. The free-energy profiles are anchored to zero in the trans state. Increasing probability density of pyramidality is color-coded from dark violet to bright yellow. Data for the other tripeptides is reported in SI Fig. A.16. (B) Free-energy profile $G$ for peptide NPA based on US (Section 3.2.3) and normalized probability density of the pyramidality $p$ (bottom; Eq. 3.21) along the $\omega$ dihedral angle. (C) Free energy of isomerization $\Delta G$ (Eq. 3.16) from $\omega$BP-REMD (replica 1 only) or US for all considered tripeptides. Error bars refer to standard deviations of subset free-energy values as detailed in Sections 3.2.5 and 3.2.5. (D) Running estimate of the isomerization free energy $\Delta G$ for tripeptides YPP, DPC and QPL from the $\omega$BP-REMD method (solid lines) and US (dashed lines). Data for the other tripeptides is reported in SI Fig. A.17. The first running free-energy estimate was taken at a simulation time of 2 ns.

angles is sampled in higher replicas (Fig. 3.3 B) and specifically, the transition state regime can be accessed, a complete free-energy profile along $\omega$ can be constructed with the WHAM method[103]. However, in the present work, the corresponding reweighting of configurations along one degree of freedom ($\omega$) yields unbiased (reweighted) probabilities ($P(\omega)$; Eq. 3.19) that are modulated by the fact that actually rigorous reweighting along one degree of freedom ($\omega$) is not possible for the $\omega$BP-REMD simulations because all four dihedral angle terms surrounding the peptide bond are biased in replicas 2-12 (Sections 3.2.2 and 3.2.5).
Both $\omega$BP-REMD analyses, i.e. based on solely replica 1 (Section 3.2.5) or based on all replicas (Section 3.2.5) yield the same PMF minima (Fig. 3.4 A, top panel; green line vs. blue dots; data for other tripeptides in SI Fig. A.16). The free-energy differences between these minima are in excellent agreement with results from US (see above and Fig. 3.4 C).
However, the free-energy values in the transition state regime show subtle differences between both methods. Maximum free energies from the $\omega$BP-REMD method, 17.3 and 17.7 kcal·mol$^{-1}$ at $\omega$ values of -93.6 and 85.0°, respectively, are slightly increased and shifted compared to corresponding values from US (14.6 and 14.9 kcal·mol$^{-1}$ at -91.4 and 101.0°, respectively). To further elucidate this discrepancy the out-of-plane deformation of the imide nitrogen atom, referred to as pyramidality $p$ (Eq. 3.21), was analyzed. This distortion is known to be an important degree of freedom in the Pro isomerization process [122]. For both methods, the pyramidality essentially vanishes when sampling the stable cis/trans-states (Fig. 3.4 A, B; bottom panel). This justifies the above-mentioned assumption that all four dihedral angles contributing to the twofold dihedral angle potential energy terms are set equal to the value of $\omega$ (Section 3.2.5) when only the calculation of the cis/trans equilibrium is of interest. Note, however, that in this case, this assumption can be entirely circumvented by only considering sampling in replica 1 (Section 3.2.5), where the twofold potential-energy contributions were left unaltered (Eq. 3.2). The pyramidality is most pronounced at $\omega$ values slightly above $\pi/2$ (mean pyramidality of 0.71 Å$^3$ for $\omega$ values in the range of $\pi/2$ up to 0.78 $\pi$, which corresponds to a distance of 0.41 Å of the nitrogen atom above the plane formed by the three bound atoms, assuming bond lengths of 1.5 Å). The omission of actual dihedral angle values for non-$\omega$ dihedral angles surrounding the peptide bond is expected to impact the PMF height. To accurately calculate transition free energy barriers, it is essential for the four distinct dihedral angle values to explicitly contribute to the configurational reweighting. For US, the pyramidality drops from 1.25 to -1.35 Å$^3$ and 1.77 to -0.95 Å$^3$ when passing the free energy maxima at $-\pi/2$ and $\pi/2$, respectively. Such extensive out-of-plane distortion of the imide nitrogen atom in the transition state is not observed in the $\omega$BP-REMD simulations,

where mean pyramidality values remain below 0.6 $\mathring{A}^3$ in the transition states. The increased out-of-plane distortion during US compared to $\omega$BP-REMD implies that the pyramidality distortion predominantly originates from the twofold cosine terms acting on the $\omega$ dihedral angle, which are significantly turned down in the higher replicas of the applied $\omega$BP-REMD method. However, in US, a substantial energetic penalty occurs at $\pm\pi/2$, where the remaining force-field terms along the three non-$\omega$ X-C$i-1$-N$i$-Y dihedral angles reach their maxima. Close to these maxima, as the pyramidality reaches its peak, signifying maximal violation of nitrogen atom planarity, and as $\omega$ increases further (beyond the transition state), the nitrogen atom abruptly shifts to the opposite side of the plane, indicated by the sign change of $p$. This observed pyramidality is rooted in the high-energy dihedral-angle force-field terms in the transition-state regime and thus of different origin and character than the orbital-based quantum-mechanical phenomenon.
Notably, the computational cost is lower for the $\omega$BP-REMD method than for US. In the latter, 24 shifted US windows were required to achieve sufficient overlap, resulting in a total simulation time of 2400 ns. In contrast, the same simulation results were obtained with 12 replicas in the $\omega$BP-REMD simulations, amounting to a total simulation time of 1200 ns.

### 3.3.2 N2 Domain of G3P

The investigation of Pro cis/trans isomerization was extended to a larger system, where the Pro of interest is embedded in an ordered, folded environment. Specifically, the N2 domain of the G3P protein, extensively studied in previous experimental double-mixing studies [120], was examined. It features a small beta sheet involving a $\beta$-hairpin loop, with the Pro of interest, Pro161, residing in the solvent-exposed loop connecting two $\beta$-strands (Fig. 3.5). Experimental cis/trans isomerization free energies are available for the wildtype and for various point mutations in the two beta strands (see Fig. 3.5 A). Again, we applied both the US and $\omega$BP-REMD method in order to calculate the isomerization free energies of the Pro of interest in the wildtype and a selection of point mutants (L125A, G158A, K163A, Y165A; Fig. 3.5 A). In the native state of the phage protein, Pro161 is in the cis isomer (see Fig. 3.5 B). Notably, Pro isomerization plays a crucial role in this protein by serving as a molecular timer that regulates the infectivity of the phage [119].

Similiar to the smaller tripeptide systems (Section 6.3.1), the $\omega$BP-REMD method was capable of inducing frequent cis/trans transitions in the wildtype of the N2

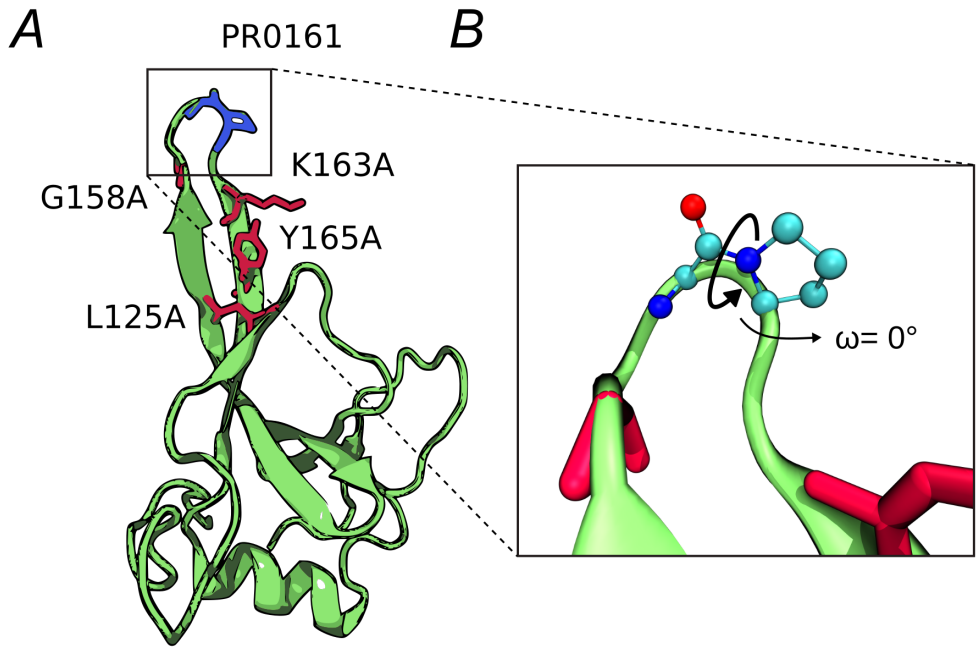



**Fig. 3.5.** N2 domain of the G3P protein. (A) Cartoon representation of the crystallized tertiary structure (PDB ID: 1G3P[138]). Pro161 is highlighted in blue stick representation and the sites of all investigated point mutations in the beta strands adjacent to Pro161 are highlighted in red stick representation. (B) Close-up view of Pro161 in the native cis isomer state.

$\beta$-hairpin within less than 1 ns in replica 1 (wildtype in Fig.3.6 A; data for mutant proteins in SI Fig. A.18 - A.22 A). As for the tripeptide systems, a potential-energy term $W_1(\omega)$ with multiplicity $n = 1$ was added along the $C_{\alpha,i-1}$-$C_{i-1}$-$N_i$-$C_{\alpha,i}$ dihedral angle (Eq. 3.2), with $V_{\omega,1} = 1.0$ kcal·mol$^{-1}$. Note that the phase of the onefold cosine term (Eq. 3.2) is shifted in comparison to the tripeptides, i.e. here $\gamma = 0$, since the cis isomer ($\omega = 0$) is more stable than the trans isomer for the N2 protein.

Mutated variants of the N2 protein reacted differently to the biasing in the $\omega$BP-REMD method. The L125A mutant experienced immediate switching to the trans state in replica 1 (SI Fig. A.21 A). In contrast, the G158A mutant adopted the trans state only after $\sim$ 15 ns simulation time (SI Fig. A.19 A). Nevertheless, both isomerization states were frequently visited in the course of 400 ns MD simulation for all systems with high exchange rates in the range of 3.3-10.0 ns$^{-1}$ (Table 3.2).

Comparing the $\omega$ dihedral angle sampling in higher replicas to the tripeptide systems a pronounced preference for dihedral angle values in the proximity of $\sim 0.6\,\pi$ is apparent for all mutations (Fig. 3.6; data for protein mutants in SI Fig. A.18 - A.22 B). This may be partially due to the onefold $W_1(\omega)$-bias potential

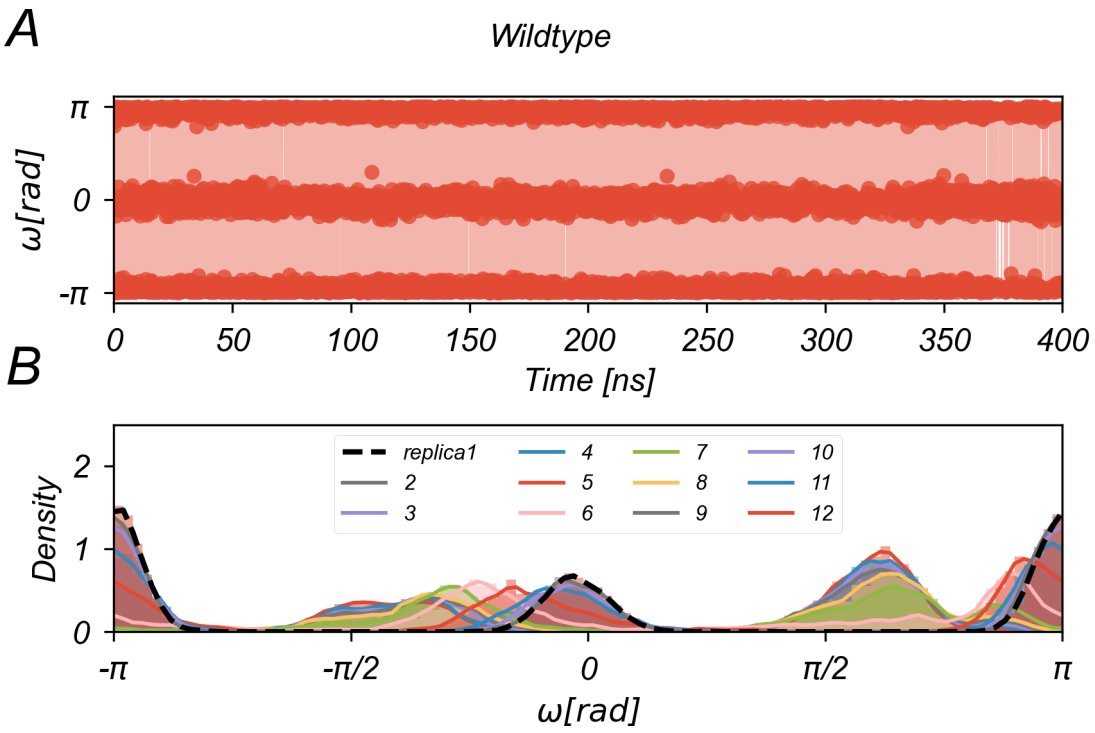



**Fig. 3.6.** Cis/trans isomerization in the N2 domain of protein G3P investigated with the $\omega$BP-REMD method (Section 3.2.2). Data for protein mutants is reported in SI Fig. A.18 - A.22. (A) Sampling of the $\omega$ dihedral angle in replica 1, where the barrier height of the dihedral angle energy profile amounts to 20 kcal·mol$^{-1}$ (Eq. 3.3), during a simulation of 400 ns length. (B) Sampling of the $\omega$ dihedral angle in all 12 replicas depicted by the probability density along $\omega$. The probability density for replica 1, presenting extensive sampling of the cis/trans states, is highlighted with the black dashed line.

disfavoring the cis state. The results demonstrate the capability of the $\omega$BP-REMD method to induce frequent cis/trans isomerization events, even if the Pro resides in a folded environment, i.e. where it is not a priori clear whether isomerization of the Pro residue may require or/and entail structural rearrangement. In the present case, the $\omega$BP-REMD method allows the isomerization to occur without any conformational changes in the vicinal $\beta$-hairpin. An important advantage of the $\omega$BP-REMD method is hence to induce a large local conformational change without perturbing the global fold of the protein. In contrast, US may lead to other structural changes in the neighborhood of the Pro residue, as evidenced here by disruptions of the hydrogen bond network between the beta strands vicinal to the Pro residue (see below).

The probability distributions sampled along the $\omega$ dihedral angle can be utilized to extract the free energy of cis/trans isomerization, as discussed in Sections 3.2.5 and 3.2.5. Across all examined variants of the N2 domain, except for the G158A mutant, the $\omega$BP-REMD method yields isomerization free energies that are in better agreement with experimental values compared to the Umbrella sampling (US) method (see Fig. 3.7 C).
Notably, the US method consistently overestimates the magnitude of the isomerization free energies. This overestimation ranges from approximately 0.26 kcal·mol$^{-1}$ for G158A to as much as 2.02 kcal·mol$^{-1}$ for the wildtype protein (see Fig. 3.7 C). The maximum deviation from experimental values is lower for $\omega$BP-REMD than for US, with the largest discrepancy observed in the G158A system, where the magnitude of the isomerization free energy is overestimated by 0.33 kcal·mol$^{-1}$. For the $\omega$BP-REMD method, the sampling based on replica 1 is sufficient to accurately predict the cis/trans equilibrium (Fig. 3.7 A; green line; data for protein mutants in SI Fig. A.23. This was already the case for the tripeptide systems (Section 6.3.1).

Comparing the free-energy profile along the $\omega$ dihedral angle to corresponding results from US, the transition state appears again slightly shifted and presents a lower free energy in the case of US (Fig. 3.7 A, B). Corresponding differences in transition rates, calculated based on transition state theory [156] and assuming a transmission coefficient of unity, are sizable and span an order of s$^{-1}$ to $10^2$ ms$^{-1}$ (SI Table A.3). The rates from the $\omega$BP-REMD method for the trans$\rightarrow$cis transition, 20.9 or 42.0 s$^{-1}$, are, however, in good agreement with experimental data (100 s$^{-1}$) [120], whereas the rates for the cis$\rightarrow$ trans transition are underestimated (2.6 or 1.3 s$^{-1}$ compared to experimentally 880 s$^{-1}$; SI Table A.3). Furthermore, in US the free-energy maximum at $\sim \pi/2$ is more favorable for all mutations, except for the G158A mutation (SI Fig. A.23). Akin to the tripeptide systems, here, too, configurations with out-of-plane distortions of the nitrogen atom in the transition

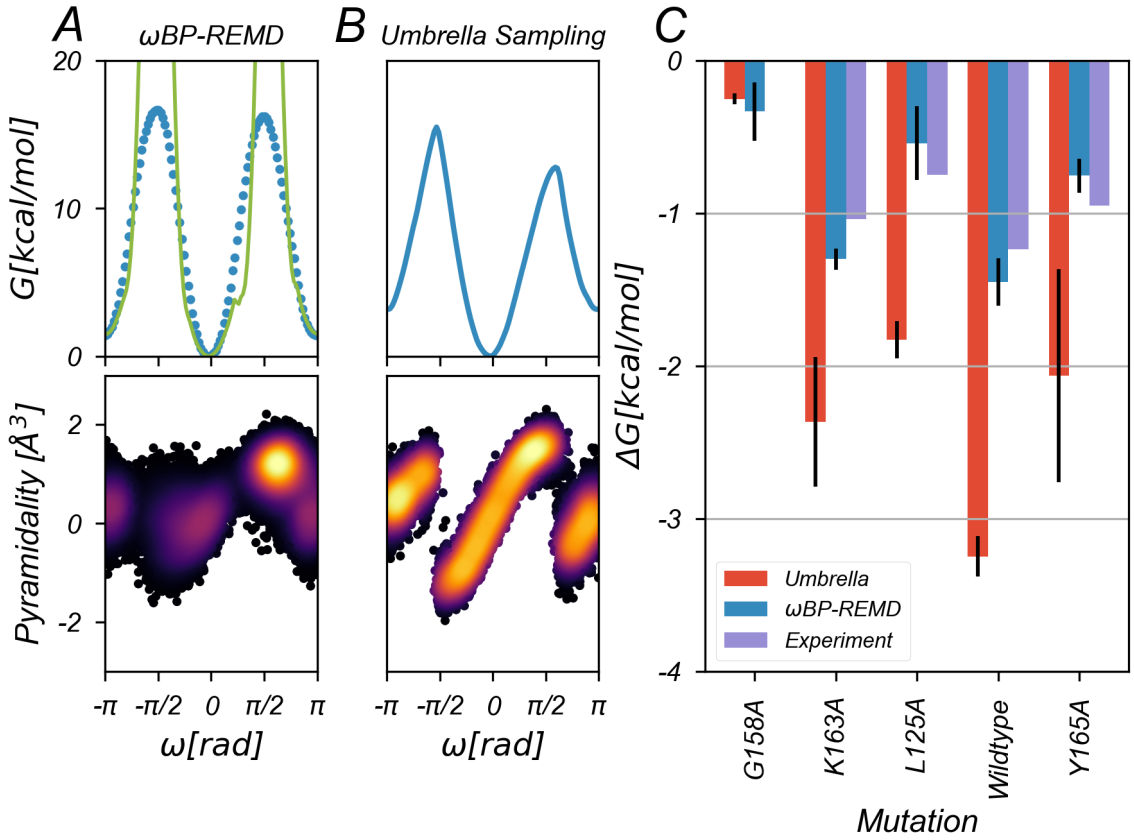



**Fig. 3.7.** Free energy of cis/trans isomerization. (A) Free-energy profile $G$ for the wildtype N2 domain of the G3P protein based on the $\omega$BP-REMD method (top) and probability density of the pyramidality $p$ (bottom; Eq. 3.21) along the $\omega$ dihedral angle. The latter is not reweighted and includes data from all replicas. The green line indicates free energies calculated from unbiased (reweighted) probabilities based on replica 1 (Eq. 4.3; Section 3.2.5), where the transition state regime is essentially unsampled. The blue dots indicate the free-energy profile calculated from unbiased (reweighted) probabilities based on all replicas (Eq. 3.19; Section 3.2.5) and the WHAM method[103]. The free-energy profiles are anchored to zero in the cis state. Increasing probability density of pyramidality is color-coded from dark violet to bright yellow. Data for protein mutants is reported in SI Fig. A.23. (B) Free-energy profile $G$ for the wildtype N2 domain of the G3P protein based on US (Section 3.2.3) and normalized probability density of the pyramidality $p$ (bottom; Eq. 3.21) along the $\omega$ dihedral angle. (C) Free energy of isomerization $\Delta G$ (Eq. 3.16) from $\omega$BP-REMD (replica 1 only) or US for the wildtype and all considered mutants of the protein. Error bars refer to standard deviations of subset free-energy values as detailed in Sections 3.2.5 and 3.2.5. Experimental isomerization free energies are from Ref. [120]

.

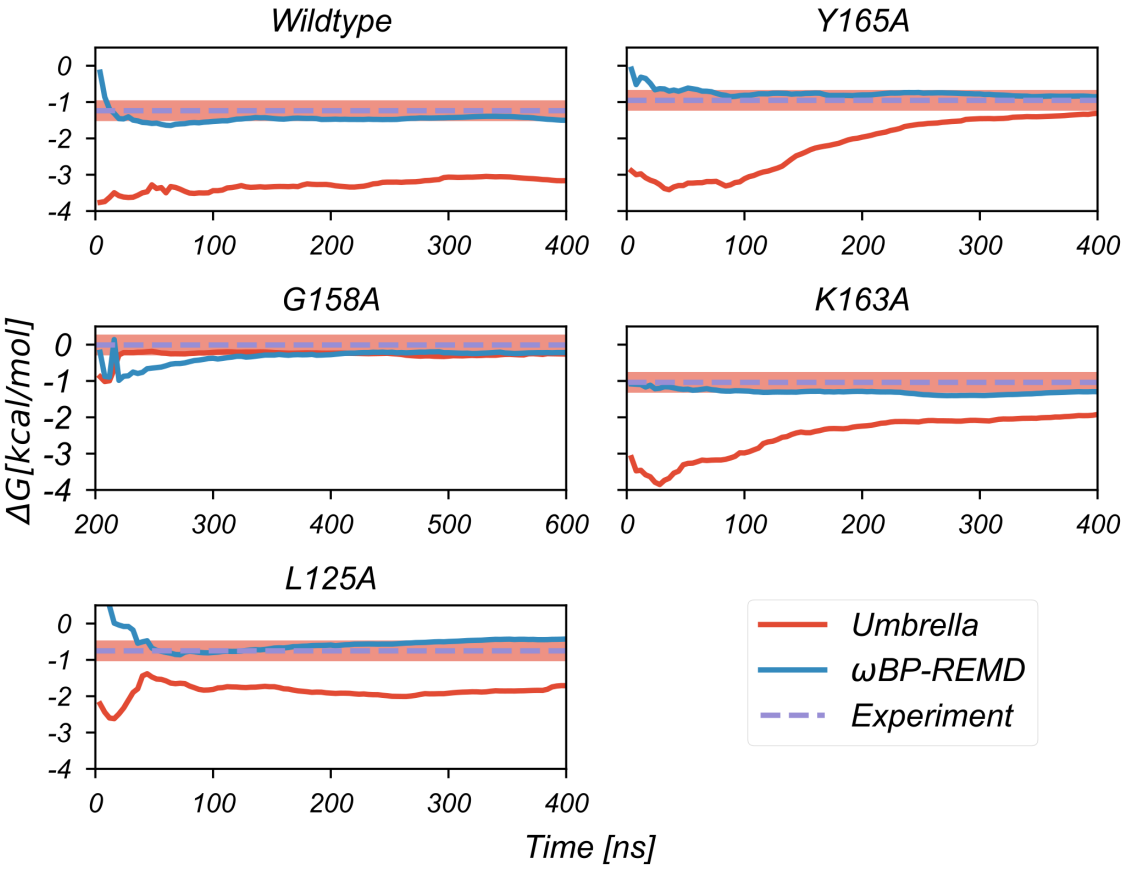



**Fig. 3.8.** Convergence behavior of the free energy of isomerization. Running estimate of the isomerization free energy $\Delta G$ (Eq. 3.16) for the wildtype protein and point mutants G158A, L125A, K163A and Y165A from the $\omega$BP-REMD method (solid blue lines) and US (solid red lines). The first running free-energy estimate was taken at a simulation time of 2 ns. An equilibration period of 200 ns was cut off from the data displayed for mutant G158A. The value of the experimental isomerization free energy[120] is indicated by the dashed blue lines and a range of $\pm k_B T$ is indicated by transparent red shading.

state are sampled in higher replicas of the $\omega$BP-REMD method. The pyramidality is highest ($\sim$ 1.8 Å$^3$) at $\omega$ values of around $0.6\pi$.

The simulations were split into subsets in order to evaluate the convergence of the isomerization free energy for both methods. Note that for the G158A mutation an equilibration period of 200 ns length was required as opposed to all other systems, where a much shorter equilibration period of 1 ns was sufficient. The longer equilibration period for the G158A mutant was chosen to allow exhaustive sampling of the backbone and the position of the introduced alanine methyl group. With the $\omega$BP-REMD method, converged isomerization free energies are reached on a short time scale of only $\sim$50 ns for all systems (Fig. 3.8, blue line) with the exception of mutant G158A, where convergence is reached after about 100 ns (after an equilibration period of 200 ns; SI Fig. A.24). The US method, on the other

hand, requires consistently up to 200 ns of sampling per window in order to give converged results.

Refolding experiments suggest stable beta sheet structures in the N2 domain of the G3P protein in both the cis and trans states of Pro161 (Fig. 3.5 A) [120]. In addition, it is argued, that the structural integrity of these beta sheets influences the cis/trans equilibria, i.e. disruption of the interactions between the beta strands adjacent to Pro161 may influence the isomerization free energy [120]. For this reason, the hydrogen bond network within the $\beta$-hairpin loop was analyzed, together with the influence of the presence or absence of the correct $\beta$-sheet hydrogen bonding pattern in the (non-native) trans state on the obtained isomerization free energies. Note, in this context, that the native cis state of the Pro allowed the sampling of stable $\beta$-hairpin structures throughout. In the trans state, however, distances between the nitrogen and carbonyl oxygen atoms involved in hydrogen bonding may lengthen more significantly beyond the threshold for a typical donor heavy atom - acceptor hydrogen bond distance of 3 Å (Fig.3.9 B). In the trans state, the first hydrogen bond of the $\beta$-hairpin between the carbonyl group of Gln157 and the NH-group of Lys163 is sampled very infrequently for both methods, with the most likely N-O distances sampled in the range of 5-10 Å(Fig. 3.9 B and C). However, for the two subsequent hydrogen bonds of the $\beta$-hairpin loop (Fig. 3.9 B) in the trans state, the $\omega$BP-REMD method samples more frequently low distances ($<$5 Å) than the US method, indicating a more stable $\beta$-hairpin loop in the trans state (Fig. 3.9 C). In fact, $\omega$BP-REMD renders distances for the second and third hydrogen bonds below this threshold in 85% or 100% of the analyzed frames of the trans state, respectively, while corresponding fractions only amount to 47% or 69%, respectively, for US (Fig. 3.9 C). Hence, for the trans state sampled with US, even distances $>$10 Å are observed rather frequently (Fig. 3.9 B and SI Fig. A.35), indicating pronounced disruption of the $\beta$-hairpin hydrogen bond network. Only the fourth hydrogen bond of the $\beta$-hairpin between the carbonyl group of Tyr165 and the NH-group of Thr156 appears stable (Fig. 3.9 and SI Fig. A.35). Strikingly, US exhibits a clear cis to trans isomerization path dependency for the structural integrity of the $\beta$-hairpin in the wildtype N2 domain with a clear disruption (i.e. lack of distances $<$5 Å) of the first hydrogen bond at values slightly above $-\pi/2$ (Fig. 3.9 B; data for protein mutants in SI Fig. A.25 - A.29). The complete lack of the first hydrogen bond in this range of $\omega$ values is not prominent in the $\omega$BP-REMD method.

Tracking the relative frequency of occurrence of the hydrogen bonds over the simulation time, it is also evident that the US windows anchored at $\pi$ sample more stable hydrogen bonds (reflected by the higher frequency of occurrence)

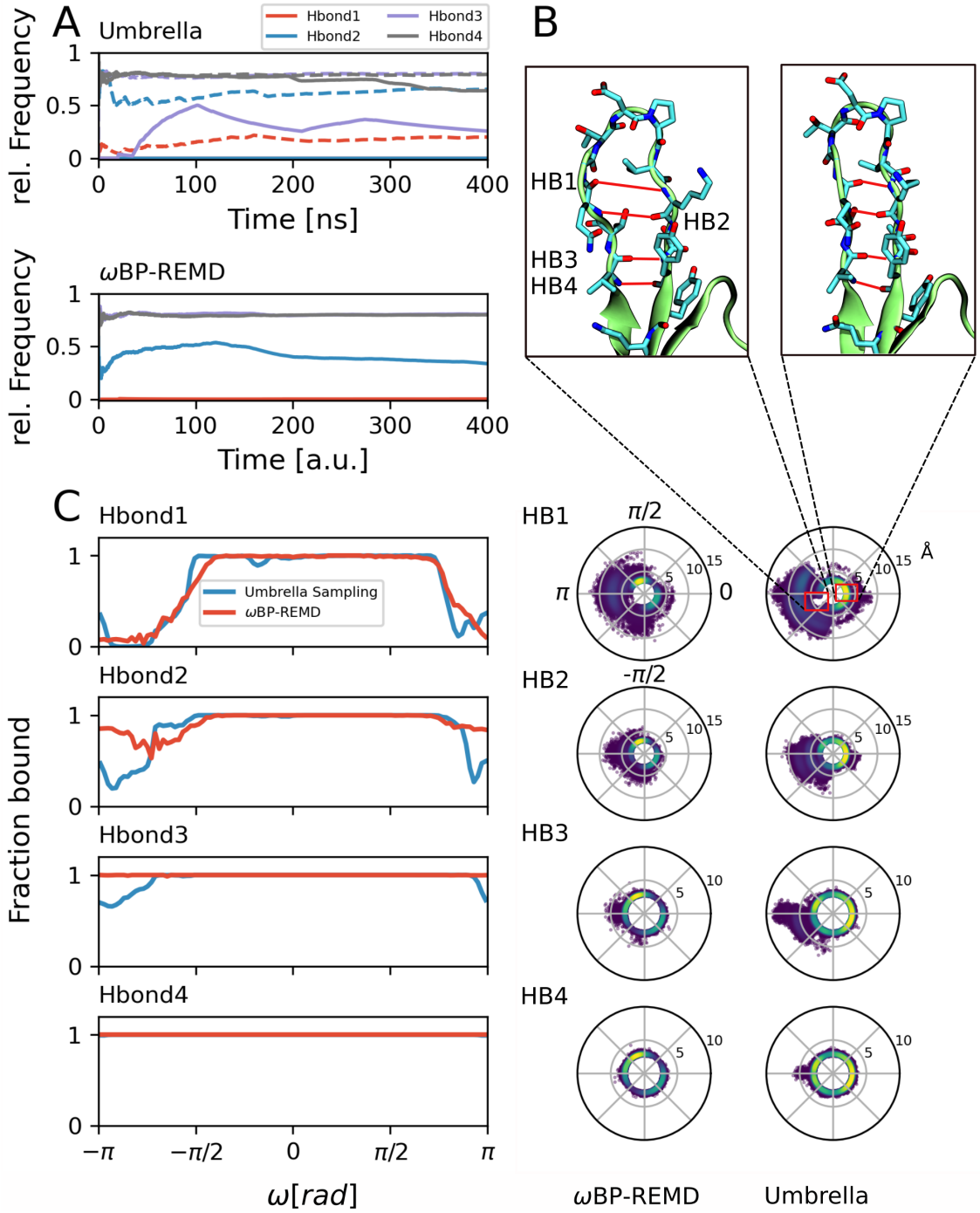



**Fig. 3.9.** Stability of inter-strand hydrogen bond interactions in the $\beta$-hairpin loop in the N2 domain of the G3P protein. (A) Running estimate of the relative frequency of hydrogen bond occurrence for hydrogen bonds 1-4 in the trans state for the wildtype N2 domain. Data for protein mutants is reported in SI Fig. A.30-A.34. Corresponding donor-acceptor distances are displayed in SI Fig. A.35. In the top panel, relative hydrogen bond frequencies for US are shown, where dashed and continuous lines pertain to the US windows at $\pi$ and $-\pi$, respectively. In the bottom panel, relative hydrogen bond frequencies for $\omega$BP-REMD in the trans state are shown. Here, time is given in arbitrary units, since the trans configurations were filtered out from the trajectory of replica 1, which samples both cis and trans states (Fig. 3.6 A), i.e. the depicted data pertains to trans configurations that are ordered in time, but not directly sequential to one another.(Caption continued on next page.)

**Fig. 3.9.** Continuation of figure caption. The number of considered frames is 6703, 8694, 7406, 9253 and 12380 from a total of $1.0{\cdot}10^4$, $1.0{\cdot}\ 10^4$, $1.0{\cdot}\ 10^4$, $1.0{\cdot}\ 10^4$ and $1.5{\cdot}\ 10^4$ frames for the wildtype protein and mutants Y165A, K163A, L125A and G158A, respectively. Note that Hbond1 and Hbond2 for the US window at $-\pi$ are essentially absent, as is Hbond1 in frames from $\omega$BP-REMD in the trans state. (B) Distances in Å of the nitrogen and carbonyl oxygen atoms constituting each respective hydrogen bond (labeled "Hbond" 1-4), plotted radially over the $\omega$ dihedral angle for the wildtype N2 domain. Data for protein mutants is reported in SI Fig. 25-29 A.25-A.29. The respective hydrogen bonds are illustrated in the top panel via red lines in a stick-representation of the $\beta$-hairpin and denote interactions between Gln157(O) and Lys163(N) for Hbond1, Gln157(N) and Lys163(O) for Hbond2, Thr156(O) and Tyr165(N) for Hbond3 and Thr156(N) and Tyr165(O) for Hbond4. Increasing probability density of distance/$\omega$-combinations is color-coded from dark violet to bright yellow. The left and right columns depict data from the $\omega$BP-REMD and US method, respectively. (C) Fraction of frames, where the distances between the nitrogen and oxygen atoms in hydrogen bonds 1-4 are smaller than 5 Å (i.e. the two atoms are considered "bound") in the wildtype N2 domain. Note that this is a looser criterion than the employed hydrogen-bond criterion (Section 3.2.5). The data is displayed for $\omega$BP-REMD and US simulations, as a function of $\omega$. Note that the bound fraction for Hbond4 is unity for both US and $\omega$BP-REMD.

as compared to the windows anchored at $-\pi$ (Fig. 3.8 A; top panel), which corresponds to the asymmetry in hydrogen bond occurrence around $\pm\pi$ observed in Fig. 3.8 B and indicates sampling of disparate $\beta$-hairpin structures in these two equivalent US windows. In general, the $\omega$BP-REMD method exhibits more stable hydrogen bond behavior over time. This can be correlated with faster $\Delta G$ convergence (Fig. 3.8). The more stable hydrogen bond behavior is also seen for the simulated protein mutants (SI Fig. A.30 - A.34). The increased structural integrity of the $\beta$-hairpin in the $\omega$BP-REMD method compared to US is an important advantage. It may be traced back to the fact that isomerization in the lowest replica is enabled via introduction of configurations from higher replicas where isomerization is possible without the occurrence of large forces due to the lower $V_{\chi[c],2}$ in the high replicas (Section 3.2.2; Eq. 3.4). In contrast, in US, isomerization is enforced by the presence of a bias potential that directly manipulates the $\omega$-dihedral angle and invokes high forces to penalize deviations from the targeted $\omega$ value (Section 3.2.3; Eq. 3.7). This also implies that the $\omega$BP-REMD method is more capable of focussing configurational sampling on the important regions of phase space, i.e. here, stable structures of the $\beta$-hairpin in both the cis and trans states. Configurational sampling according to low Boltzmann factors in the states of interest in turn increases the computational efficiency of the method. This is an important advantage in comparison to other advanced sampling schemes like e.g. metadynamics [135, 139] or (Gaussian) accelerated molecular dynamics [131, 132], where transitions between states are enabled by smoothening the

potential-energy surface, i.e. local minima of interest may eventually be no more preferentially accessed (Fig. 3.2 A; top panel).

## 3.4 Conclusion

The primary aim of this study was to efficiently and accurately compute Pro isomerization free energies using classical atomistic MD simulation. To achieve this, the $\omega$BP-REMD method was developed with two key features: ($i$) decreasing barrier heights of the energy profile along the replica ladder for the concerned peptide bond dihedral angle and ($ii$) an additional onefold bias potential to destabilize a specific Pro isomer (cis or trans), promoting balanced sampling of the two isomer states.

The first feature enhances sampling in the lowest replica through the exchange of configurations, while the second feature dictates computational efficiency by achieving balanced sampling of the cis and trans isomer states. The $\omega$BP-REMD method focuses on important regions of phase space in the lowest replica, allowing rapid convergence of the sampled cis/trans equilibrium. Cis/trans exchange rates in the lowest replica ranged from 3-11 $ns^{-1}$ for all investigated systems, even when Pro was embedded in a folded protein structure. This ensured excellent transition statistics while maintaining the native protein fold.

To correct for the potential-energy bias in replica 1, reweighting of the sampled probability distribution along the $\omega$ dihedral angle was employed. This correction enabled the accurate quantification of the cis/trans equilibrium, though it did not permit estimation of the isomerization barrier height due to the absence of transition state regime access in replica 1. To address this limitation, the Weighted Histogram Analysis Method (WHAM) was utilized to unbias and combine sampling across the entire replica ladder, constructing a complete free-energy profile along the $\omega$ dihedral angle.

Comparisons between the $\omega$BP-REMD and standard US methods were conducted using solvated tripeptides and the N2 domain of the G3P protein. While both methods exhibited comparable performance for tripeptide systems with excellent agreement of calculated isomerization free energies, the $\omega$BP-REMD method outperformed US in case of the protein. Specifically, $\omega$BP-REMD demonstrated better

agreement with experimental isomerization free energies (maximum absolute deviation of 0.3 kcal·mol$^{-1}$ for the $\omega$BP-REMD method vs. 2.0 kcal·mol$^{-1}$ for US), faster convergence (approximately 50 ns vs. 200 ns), and maintained the $\beta$-hairpin structure, contrary to disruptions observed in US simulations. Experimental data[120], however, suggests a stable $\beta$-sheet structure in the cis and trans state. We trace back the increased structural integrity of the $\beta$-hairpin in the $\omega$BP-REMD method to the fact that enhanced sampling in this method is introduced in a gentle way through canonical exchange of configurations from higher replicas where the transition states of isomerization can be accessed without severe energetic penalties. In contrast, during US, high forces associated with the harmonic bias potential enforce the sampling of desired dihedral angle values around the peptide bond and configurational sampling may thus be driven towards physically less relevant phase-space regions.

Differences in transition-state thermodynamics and structure between the $\omega$BP-REMD and US methods were observed in our study. The $\omega$BP-REMD method consistently yielded transition states that were more unfavorable, with discrepancies of up to about 3 kcal·mol$^{-1}$ compared to the US method. Additionally, slight variations were noted in the $\omega$ dihedral angle value and nitrogen atom pyramidality.

The distinctive reaction coordinates employed by both methods contributed to the differences in the transition-state free energy. It remains uncertain to what extent these variations arise from the chosen reaction coordinates versus insufficient sampling in the transition-state region. It is likely that the former plays a non-negligible role because the differences are observed even for small tripeptide systems where sampling issues are presumed negligible on the simulated timescale.

In the US method, the reaction coordinate for the free-energy profile is simply the $\omega$ dihedral angle. In contrast, the $\omega$BP-REMD method utilizes sampled probability distributions that depend not only on the value of $\omega$ but also on the three other dihedral angles surrounding the peptide bond. Due to these different descriptions of the reaction coordinate, the free-energy of the transition-state regime can differ for two main reasons.

The first reason involves an artifact in the present reweighting formalism, where the values of the three non-$\omega$ dihedral angles are not explicitly considered in the reweighting but are set equal to $\omega$ for simplicity. This approximation may not be justified when significant out-of-plane distortions of the carbonyl or NH-groups occur, as in the transition-state regime. Characterization of the isomerization

barrier height (i.e. kinetics) might necessitate small amendments to the reweighting procedure.

The second reason for lower transition-state free energies in the US method compared to the $\omega$BP-REMD method could be the result of biasing only one dihedral angle (rather than the four involved in the peptide bond). The transition state in the US method may experience stabilization through relaxation of the other dihedral angles to adapt optimally to the umbrella restraint. Indications of this are seen in the significantly different pyramidalities of the nitrogen atom sampled by the two methods in the transition state regime, where the US method generally exhibits higher nitrogen atom pyramidalities.

It's crucial to note that the description of the transition state with both methods may still be unrealistic, as it does not truly reflect the quantum-mechanical characteristics of the nitrogen atom's out-of-plane distortion. Instead, it reflects the influence of classical molecular-mechanics force-field terms and potentially additional restraints. In conclusion, these discussions about the transition state are provided for completeness and do not impact the obtained cis/trans equilibria, which were the primary focus of the study.

We think that the proposed $\omega$BP-REMD method has the potential to allow the efficient and accurate calculation of Pro isomerization free energies also in larger protein environments, in particular, also in systems where multiple Pro residues occur and Pro isomerization states may depend on each other [130, 157]. Correlated changes in Pro isomerization states in proteins can be investigated by simultaneous application of bias potentials for all Pro residues of interest and analysis of the resulting multidimensional probability distributions. With US, such a study would require stratified sampling along multiple reaction coordinates which entails a tremendous increase in compute and sampling effort.

Characterizing Pro residues based on their free energy of isomerization and exploring potential correlations between isomerization states represent crucial steps toward comprehending the role of Pro residues in proteins. These aspects will be further investigated in the upcoming chapter.

# 4

# Conformational Coupling of Proline Cis-Trans Isomerization in Bradykinin

[1]

> *In nature we never see anything isolated, but everything in connection with something else which is before it, beside it, under it, and over it.*
>
> **— Johann Wolfgang von Goethe**

It is well known that proline (Pro) cis-trans isomerization plays a decisive role for folding and stabilization of proteins. However, the conformational coupling between the isomerization states of different Pro residues in proteins during conformational adaptation processes remains less explored.

This study aims to investigate the coupled cis-trans isomerization of three Pro residues using bradykinin (BK), a partially unstructured nonapeptide hormone, as a model system. $\omega$BP-REMD (details in Section 3) allows us to thoroughly sample all combinations of Pro isomer states and obtain converged probability densities for all eight state combinations within 885 ns $\omega$BP-REMD simulations.

In agreement with experiment the all-trans state is seen to be the preferred isomer of zwitterionic aqueous BK. In about a third of its structures, this state presents the characteristic C-terminal $\beta$-turn conformation, however, other isomer combinations also contribute significantly to the structural ensemble. Unbiased probabilities were projected onto the peptide bond dihedral angles of the three Pro residues, revealing the interdependence of individual Pro isomerization states, indicating a potential coupling of different Pro isomers.

Remarkably, the cis/trans equilibrium of a Pro residue can shift by up to 2.5

[1] This chapter has been published in similar form [158].

kcal·mol$^{-1}$ depending on the isomerization state of other Pro residues. For instance, simulations suggest that the cis state of Pro7 becomes favored compared to its trans state when Pro2 is switched from trans to cis.
These findings highlight the effectiveness of the $\omega$BP-REMD methodology and propose that the coupling of Pro isomerization states may play an even more crucial role in larger folded proteins subjected to additional conformational restraints.

## 4.1 Introduction

Proline (Pro), one of the 20 natural amino acids, stands out due to its unique characteristics. Its cyclic side chain connects the $C_\alpha$ and amide nitrogen atom, making the cis and trans states of a peptide bond preceding the Pro residue (known as a prolyl peptide bond) more sterically similar compared to other amino acids. Sometimes, the cis isomer of a prolyl peptide bond is only about 0.5 kcal·mol$^{-1}$ less stable than the trans isomer [159], contributing to the presence of approximately 7% of prolyl peptide bonds in the cis state [107]. In contrast, non-prolyl peptide bonds strongly prefer the trans state. The ability of prolyl peptide bonds to adopt both isomerization states, coupled with their slow cis/trans isomerization rates (on the order of seconds to minutes), underscores their significance in biological functions. Certain Pro residues act as molecular timers [119] and significantly influence protein folding [109]. It has been shown before that Pro isomerization and conformational changes such as protein folding may be coupled [160]. For example, in protein folding processes, the reversible work required to isomerize a Pro residue to its native state can be offset by the free energy gained when the protein achieves a stable fold. Such a coupling was investigated by unfolding and refolding experiments of the N-terminal domain of the gene-3 protein of the phage fd [120, 160]. In the previous Chapter 3, we performed extensive molecular dynamics simulations to calculate Pro isomerization free energies in the wildtype and various mutant systems of this protein as well as corresponding unfolded-state models. Here, we perform Pro isomerization free energy calculations in a peptide where multiple Pro residues occur, bradykinin (BK).

In 1981, Levitt conducted a study on a folded protein with multiple Pro residues, bovine pancreatic trypsin inhibitor. In this investigation, Levitt calculated energy differences associated with Pro isomerization states [117]. The findings revealed notable variations in cis/trans energy differences among the Pro residues, and the magnitude of these differences correlated with the experimentally observed

impact of the "wrong" (non-native) isomerization state on the proper folding of the protein. For instance, Pro13 exhibited a very low energy difference between the cis and trans states, suggesting its capability to freely isomerize even within the folded protein. On the other hand, Pro8 displayed a considerably high energy difference between the cis and trans states, indicating its potential to impede the folding process to the native state when in the "wrong" cis form.

Nevertheless, in scenarios lacking structural constraints, such as in short peptides or disordered proteins, it is generally believed that there is no coupling between Pro isomerization states and biomolecular conformation [161, 162].

In the present work, we amend this view. For the partially unstructured nonapeptide BK, we illustrate, using $\omega$BP-REMD [105] that distinct cis/trans isomerization states of the three Pro residues in BK may sample distinct conformational clusters. We also show that the distinct cis/trans isomerization states of the three Pro residues may occur in a non-independent fashion, i.e. the cis/trans equilibrium of a given Pro residue may depend on the isomerization states of other Pro residues. We refer to this phenomenon as *coupling* of proline cis-trans isomerization states. It was observed before using ion mobility-mass spectrometry that BK conformation depends on the isomerization states of the Pro residues [163]. Furthermore, Yang et al. performed molecular dynamics (MD) simulations to sample the eight possible isomerization states of BK and noted a possible mutual dependence of the isomerization states of the individual Pro residues [164]. However, the study presented in this chapter is, to our knowledge, the first work to provide detailed and complete insight into the coupling of all the isomerization states of the Pro residues in BK by systematically quantifying the dependence of the cis/trans equilibrium of a given Pro residue on the isomerization states of the remaining Pro residues. We refer to these isomerization free energies as *conditional* isomerization free energies. BK, besides being an ideal system to study the connection between Pro isomerization states and molecular conformation and/or the isomerization equilibria of the other Pro residues, is a very interesting system in structural biology because it features both ordered and disordered regions which prevents a complete determination of its structure.

Numerous experimental efforts have been made to determine BK structure in solution. Circular dichroism,[165–167] Raman spectroscopy,[168] nuclear magnetic resonance (NMR)[167, 169–171] and molecular modeling studies [172–174] reveal that the BK Ser6-Arg9 residues are able to adopt a $\beta$-turn motif. The preference for the $\beta$-turn motif is enhanced by aprotic solvents, apolar media or micelles [170]. Thus, the extent of conformational structure in BK depends crucially on the envi-

ronment. It is mainly due to the N-terminal Arg1-Phe5 region that BK is described as a partially unstructured peptide. This region appears to be a random-coil like segment with no dominant structural features. Ion mobility and mass spectrometry measurements in the gas phase provide evidence for the coexistance of different conformations in the Arg1-Phe5 region, depending on protonation states and solvent [175]. The impact of protonation state on the structure of BK was also observed in computational studies [176, 177]. Pro residues are key in establishing the diversity of conformations of free BK [163]. Specifically, distinct combinations of cis or trans forms of the three residues (Pro2, Pro3, Pro7) are responsible for some of the populations observed experimentally. Interestingly, conformers of $[BK+3H]^{3+}$ (meaning the peptide with neutral C-terminus while the two arginine residues and the N-terminus are positively charged) in solution were seen to have a surprisingly high preponderance to incorporate cis Pro configurations [163].

Concerning biological function, BK appears an extraordinary versatile peptide. It possesses an astonishing ability to wield diverse roles in regulating numerous physiological processes. It can bind to the endothelial G-protein-coupled receptors (GPCRs) $B_1$ and $B_2$ and thus initiate signal transduction to exert potent pharmacological and physiological effects ranging from blood pressure regulation, vasodilation or pain response to inflammation[178–180]. BK has also been shown to exert potent antithrombogenic, antiproliferative and antifibrogenic effects [181] which may be exploited towards clinical benefit. According to NMR spectroscopic studies, binding of BK to the receptor $B_2$ occurs in a relatively elongated conformation, with all three Pro residues in the trans conformation and a $\beta$-turn at the C-terminus [182, 183]. Hence, interestingly, only a single BK conformer appears to be found in complex with the receptor. It is still largely unclear how other, "inactive" BK conformers may convert into conformers capable of receptor binding and whether there are other, still undiscovered receptors available for binding these conformers. Clearly, complete characterization of BK conformations in different environments is important to understand its biological role and may aid in the design of more effective receptor agonist and antagonist analogues.

Most computational models for BK presume an all-trans state. Note in this context that, in the present work, we distinguish the isomerization state of a given Pro residue from the isomerization state of the peptide BK. The former can exist in two forms (cis or trans), whereas the latter can exist in eight possible forms (each Pro residue being present in the cis or trans state). Accurate computational models have to incorporate the possibility of Pro cis/trans isomerization and the presumption of a Pro residue in the trans state might not be justified, not only for BK but for smaller peptide systems in general. Frequently found Pro residues in neuropeptides might

also indicate a biological relevance of Pro cis/trans isomerization [184]. However, any investigation of Pro/cis trans isomerization with explicit-solvent atomistic MD simulation is hampered by the extremely long timescale on which the isomerization occurs. Hence, enhanced sampling methods have to be used to make the process accessible on the nano- or millisecond timescale. For example, Yang et al. [164] used a combination of metadynamics and integrated tempering sampling to sample the eight different isomerization states of $[BK+3H]^{3+}$ in explicit solvent in 1 $\mu$s trajectories. The all-trans state was found to be the preferred state. This contrasts with the mass spectrometry-based study of Pierson et al. [163], where aqueous $[BK+3H]^{3+}$ was found to prefer a state with Pro2, Pro3 in the trans and Pro7 in the cis isomer. Note that the study involved three biased reaction coordinates, i.e. the three Pro-preceding peptide bond dihedral angles. Sampling thus ultimately occurs on a three-dimensional flattened free-energy surface which implies a tremendous compute effort to adequately capture the phase-space areas of interest. Multidimensional free-energy surfaces can in principle be calculated in a more compute-efficient manner using exchanges between the multiple bias potentials [185] such as in parallel-bias metadynamics [186]. However, the computational efficiency of this method, especically the filling-up rate, is very parameter-sensitive. In addition, diffusion through the high-dimensional reaction-coordinate space may be slow such that it takes long to adequately sample all states of interest. The latter problem may, however, be overcome with selective temperature increases of fictitious degrees of freedom coupled to the reaction-coordinate variables [187].

In the present study we use $\omega$BP-REMD [105], to quantitatively investigate coupled Pro isomerization processes in aqueous, zwitterionic BK. Our approach, based on Hamiltonian replica-exchange simulations, allows simultaneous cis/trans exchanges of all three Pro residues in BK. By equally focussing the sampling in the lowest replica on all the physically important regions of phase space, which is achieved through an additional bias potential on each of the three prolyl peptide bond dihedral angles, this method can accurately capture rarely occurring cis/trans isomerization states and reveal potential dependencies between the different isomerization states. There are two crucial differences in comparison to metadynamics which entail the excellent computational efficiency of our method: First, sampling in the lowermost replica does not occur on a flattened free-energy surface. This means that the sampling in this replica is strongly focussed on the interesting phase-space areas, namely the cis and trans states of each prolyl peptide bond dihedral angle. Second, rather than having to sample the volume of the three-dimensional conformational space spanned by the three combined reaction coordinates as in metadynamics, this method is based on independent biasing of

the three reaction coordinates. This implies a great reduction in compute effort compared to a standard metadynamics approach.

The main objective of the present work was to apply $\omega$BP-REMD simulations to BK to enhance isomerization events of its three Pro residues and thus elucidate the interdependence of the three cis/trans equilibria. Our simulations of zwitterionic BK in water reveal that:

1. Distinct conformational clusters are sampled by the different isomerization states of BK.

2. The all-trans state is the preferred state and in about one third of the simulation frames it adopts a C-terminal $\beta$-turn motif presenting the characteristic Ser6-Arg9 hydrogen bond.

3. Isomerization of the three Pro residues can occur in a coupled fashion, i.e. the isomerization state of a given Pro residue can influence the cis/trans equilibrium of other Pro residues.

4. The influence of Pro isomerization states on the cis/trans equilibrium of other Pro residues may be large, i.e. conditional isomerization free energies may differ by up to 2.5 kcal·mol$^{-1}$.

Notably, to our knowledge, this is the first time that a strong coupling of Pro isomerization states has been illustrated and quantified via conditional isomerization free energies. In addition, it is noteworthy that such coupling is observed in a system like BK, i.e. a short, flexible peptide chain devoid of significant conformational restraints. This suggests that the coupling of Pro isomerization states may be even more pronounced in systems possessing more conformational restraints as e.g. represented by defined secondary structure elements in folded proteins.

## 4.2 Methods

All MD simulations were performed with the CUDA-accelerated PMEMD version of the AMBER20 software suite[147]. For the peptide, the ff14SB force field[141] was used. Ions were described with the parameters of Joung & Cheatham [188]. Water was represented explicitly with the TIP3P water model [143].

### 4.2.1 $\omega$BP-REMD: Simulation Setup and Details

The sequence of the peptide BK is RPPGFSPFR. The crystal structure of $B_2$-receptor-bound BK (PDB ID: 6f3w) [189] served as starting structure for MD simulations. The N- and C-termini of the peptide were charged $NH_3^+$- and $COO^-$-groups, respectively and the two arginine residues were protonated, resulting in a net charge of +2e. This charge state is used for all simulations in the present study. The peptide was solvated in an octahedral box with a minimum distance of 10 Å to the box boundaries. Sodium and chloride ions were added to neutralize the system and reach an ion concentration of 100 mM. After an energy minimization of 2000 steps of steepest descent, the system was gradually heated up to 300 K while keeping positional restraints on the peptide non-hydrogen atoms. The restraints were gradually released during another simulation of 1 ns length, followed by an unrestrained equilibration simulation of 20 ns at a constant pressure of 1 bar. Temperature and pressure were maintained at the target values by weak coupling, [84] using coupling times of 0.1 and 2.5 ps, respectively. A non-bonded cutoff of 9 Å was used to truncate Lennard-Jones interactions and real-space interactions in the PME method [77]. All bonds were kept at their optimal lengths using the SHAKE algorithm [82]. Newton's equation of motion was integrated numerically with the Velocity-Verlet algorithm [190]. The hydrogen mass repartitioning scheme was used, allowing a time step of 4 fs [83].

Based on the equilibrated structure, $\omega$BP-REMD simulations according to our previous work [105] were started. The setup included 12 replicas differing in the force constant $V_{\chi,2}$ of the dihedral angle potentials for the dihedral angles surrounding the prolyl peptide bonds of each of the three Pro residues of BK (Pro2, Pro3 and Pro7). Equidistant spacings from $V_{\chi,2} = 0$ kcal·mol$^{-1}$ in replica 12 up to $V_{\chi,2} = 2.5$ kcal·mol$^{-1}$ in replica 1 were chosen for each of the four torsion-energy terms pertaining to the four dihedral angles surrounding each prolyl peptide bond, resulting in the effective potentials shown in Fig. 4.1 B. This means that unhindered Pro cis/trans transitions are possible in replica 12, whereas replica 1 represents the physical dihedral angle potentials with barrier heights of 20 kcal·mol$^{-1}$, i.e.

$$E(\omega) = \sum_{i=1}^{4} V_{\chi,2} \left[1 + \cos(2\Phi_i + \gamma_{\Phi_i,2})\right] \quad , \tag{4.1}$$

where $\omega$ denotes the respective prolyl peptide bond angle ($C_{\alpha,i-1}$-$C_{i-1}$-$N_i$-$C_{\alpha,i}$; for residue $i$ referring to Pro2, Pro3 or Pro7), the sum runs over all four $\Phi_i$ denoting the

four involved dihedral angles ($C_{\alpha,i-1}$-$C_{i-1}$-$N_i$-$C_{\alpha,i}$, $O_{i-1}$-$C_{i-1}$-$N_i$-$C_{\alpha,i}$, $O_{i-1}$-$C_{i-1}$-$N_i$-$H_i$ or $C_{\alpha,i-1}$-$C_{i-1}$-$N_i$-$H_i$) and $\gamma_{\Phi_i,2} = 180°$ is the phase shift.

By exchanges between replicas according to the Metropolis criterion, [90] the respective configurations propagate into replica 1 and enhance the sampling of cis/trans isomerization in this replica. To achieve a balanced sampling of the cis/trans isomers of each Pro residue, a small potential-energy term with a cosine form of multiplicity $n = 1$, force constant $V_{\omega,1} = 1$ kcal·mol$^{-1}$ and phase shift $\gamma_{\omega,1} = 180°$,

$$W_1(\omega) = V_{\omega,1}\left[1 + \cos(\omega + \gamma_{\omega,1})\right] \tag{4.2}$$

was added in all replicas to adjust the cis/trans equilibria and destabilize the otherwise favored all-trans state of BK (Fig. 4.1 B, red line). Each replica was simulated for 885 ns, with exchange attempts [90] between neighboring replicas occurring every 250 steps. Acceptance rates between neighboring replicas were in the range of 28-89%. During the $\omega$BP-REMD simulations, coordinates were written to file every 40 ps for subsequent analysis.

### 4.2.2 Free-Energy Calculation

Eight possible cis/trans isomerization states, involving the three Pro residues Pro2, Pro3 and Pro7, can be defined for BK. Probabilities for BK to adopt any of these states were calculated from the sampling observed in the trajectory of replica 1. The three-dimensional space describing the three Pro $\omega$ dihedral angles ($\omega_2$, $\omega_3$, $\omega_7$), each ranging from -180 to 180°, was binned into cubes with 50 bins along each axis to obtain the biased probability density $\tilde{P}(\omega_2, \omega_3, \omega_7)$. The probability is biased because the physical Hamiltonian in replica 1 was augmented by the additional potential energy contribution given in Eq. 4.2. $\tilde{P}$ was then reweighted via [34]

$$P(\omega_2, \omega_3, \omega_7) = \exp^{-\beta(G_1 - G_0)} \exp^{\beta[W_1(\omega_2)+W_1(\omega_3)+W_1(\omega_7)]} \tilde{P}(\omega_2, \omega_3, \omega_7) \quad , \tag{4.3}$$

where $\beta = (k_B T)^{-1}$, $k_B$ is Boltzmann's constant, $T$ the absolute temperature, $G_1$ is, to within an additive constant, the free energy associated with the biased sampling in replica 1,

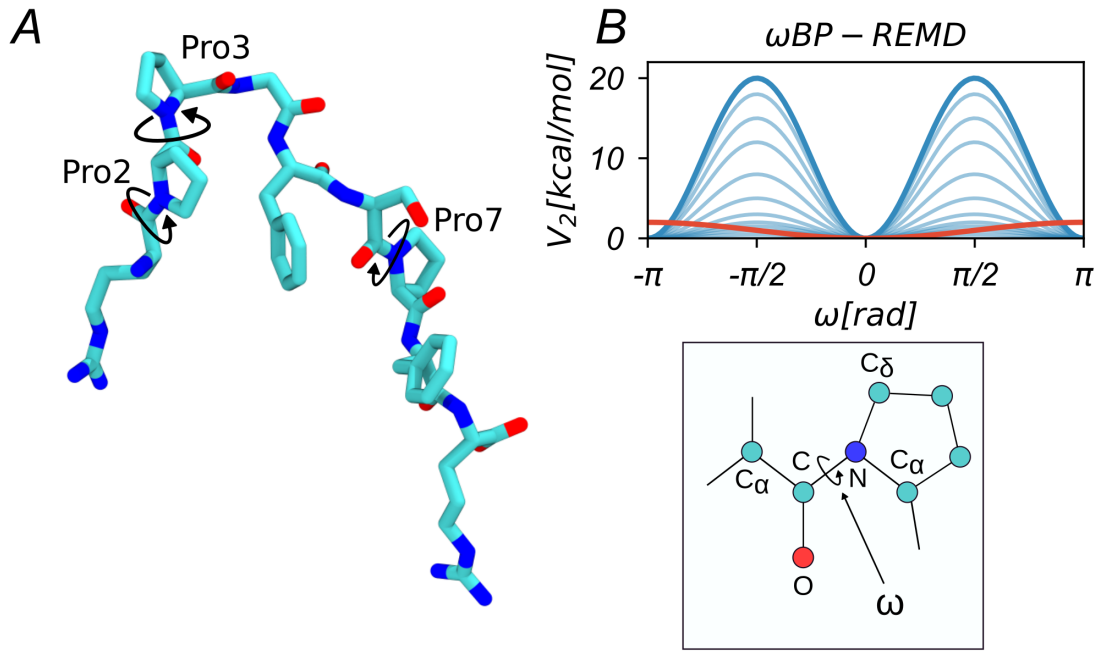



**Fig. 4.1.** (A) The zwitterionic nonapeptide BK with Pro residues at positions 2, 3 and 7, shown in stick representation in an arbitrary conformation observed during the performed MD simulations. Torsions around the respective prolyl peptide bonds are highlighted. (B) The $\omega$ dihedral angle of Pro residue $i$ is given by atoms $C\alpha_{i-1}$-$C_{i-1}$-$C\alpha_i$-$N_i$ (bottom). $\omega$BP-REMD simulations are done with simultaneous modification of the dihedral angle potentials of all three prolyl peptide bonds. Replicas 1 to 12 have Hamiltonians with decreasing barrier height of Pro cis/trans isomerization (blue lines). The Hamiltonian for replica 1, while containing the physical dihedral angle energy profile with a barrier height of 20 kcal·mol$^{-1}$ (Eq. 4.1), has an additional one-fold cosine term (Eq. 4.2 ; red line) that penalizes the trans state of all Pro residues. This one-fold cosine term is also present in the Hamiltonian of the other replicas (top).

$$G_1 = -k_B T \ln \left[ \int d\mathbf{r}^N \exp^{-\beta[U(\mathbf{r}^N)+W_1(\omega_2)+W_1(\omega_3)+W_1(\omega_7)]} \right] \quad , \tag{4.4}$$

and $G_0$ relates to the free energy associated with the sampling in the unbiased system,

$$G_0 = -k_B T \ln \left[ \int d\mathbf{r}^N \exp^{-\beta U(\mathbf{r}^N)} \right] \quad , \tag{4.5}$$

where the integration is done over the $3N$-dimensional vector $\mathbf{r}^N$ containing the coordinates of all atoms and $U(\mathbf{r}^N)$ is the physical potential energy of the system. Note that kinetic energy contributions are omitted throughout because here only changes in configurational variables are considered and that pressure-volume contributions are neglected due to their generally small magnitude. Proper normalization of $P$ in Eq. 4.3 was ensured by the condition

$$\iiint d\omega_2 \, d\omega_3 \, d\omega_7 \, P(\omega_2, \omega_3, \omega_7) = 1 \quad . \tag{4.6}$$

Probabilites for each isomerization state were calculated by integrating the probability density over the corresponding $\omega$ values. The integration was done numerically using Simpson's rule and the dihedral angle range of the trans state was defined as $180 \pm 90°$ and that of the cis state as $0 \pm 90°$. This coarse criterion is adequate to define the isomerization states in replica 1 because cis/trans states are well separated in replica 1 at $\pm 90°$ (Figs. 4.2 B, 4.4 and 4.5). Free energies were computed via Boltzmann inversion,

$$G(state) = -k_B T \ln(P(state)) \quad , \tag{4.7}$$

where "$state$" refers to the triple $\omega_2, \omega_3, \omega_7$ residing in one of the possible eight isomerization states of BK, using the three-letter terminology described in Section 4.1. Two- or one-dimensional probability densities and corresponding free-energy surfaces were obtained in a similar fashion by projection,

$$P(\omega_i, \omega_j) = \int d\omega_k \, P(\omega_i, \omega_j, \omega_k) \quad , \tag{4.8}$$

and

$$P(\omega_i) = \iint d\omega_j \, d\omega_k \, P(\omega_i, \omega_j, \omega_k) \quad , \tag{4.9}$$

where $i$, $j$ and $k$ refer to either of the three Pro residues, Pro2, Pro3 and Pro7.

### 4.2.3 Conditional Isomerization Free Energies

Based on the one-dimensional probability densities (Eq. 4.9) for a Pro residue $i$, the potential of mean force (PMF) along the concerned $\omega$ dihedral angle reaction coordinate can in principle be obtained by Boltzmann inversion,

$$G(\omega_i) = -k_B T \ln(P(\omega_i)) \quad . \tag{4.10}$$

The isomerization free energy for a cis-to-trans transition of Pro residue $i$,

$$\Delta\tilde{G}_{\text{cis}\rightarrow\text{trans}}(i) = G(\omega_i = 180°) - G(\omega_i = 0°) \quad , \tag{4.11}$$

refers to an ensemble average over the isomerization states of the remaining Pro residues $j$ and $k$. Here, we define the conditional isomerization free energy

$$\Delta G_{\text{cis}\rightarrow\text{trans}}(i; s_j, s_k) = G(\omega_i = 180°; \omega_j = s_j, \omega_k = s_k) - G(\omega_i = 0°; \omega_j = s_j, \omega_k = s_k) \tag{4.12}$$

as the isomerization free energy of Pro residue $i$ with the remaining Pro residues $j$ and $k$ residing in defined isomerization states $s_j$ and $s_k$, respectively. This allows a detailed and systematic analysis of the influence of the isomerization state of different Pro residues on the isomerization equilibrium of another Pro residue of interest. Note, that in practice, ranges of $\omega$ dihedral angle values of $\pm 90°$ around the values of 0 and $180°$ were used to define a certain state.

### 4.2.4 Further Analyses

The sampling of isomerization events was analyzed by the running mean of isomerization states over 100 consecutive simulation frames,

$$\overline{\Omega}(t_N) = \frac{1}{100} \sum_{i=N-50}^{N+50} \Omega(t_i) \quad , \tag{4.13}$$

where variable $t_i$ denotes a particular simulation time point, $\Omega(t_i) = 1$ if $\omega(t_i)$ of the considered Pro residues pertains to the trans state and $\Omega(t_i) = 0$ if $\omega(t_i)$ pertains to the cis state. Since frames were written to file every 40 ps, a value of $\overline{\Omega}(t_N)$ close to 0.5 indicates balanced sampling of cis and trans states in a 4 ns time interval centered around $t_N$.

Henceforth we use a three letter representation of BK isomerization states, where letters "T" and "C" refer to the trans and cis isomers of a given Pro residue. The first letter represents the isomerization state of Pro2, the second letter that of Pro3 and the third letter that of Pro7. For instance, the all-trans isomer of BK is then denoted as TTT.

A clustering of the sampled configurations in replica 1 was done by hierarchical density-based clustering using the HDBSCAN module of scikit-learn [191]. This clustering was done in a combined fashion, i.e. on the entire trajectory of replica 1 involving all possible cis/trans isomerization states. Input features were selected by using hydrogen-bond contacts best suited for discrimination between the conformations sampled by different isomerization states. This was done by first identifying all formed hydrogen-bond contacts during the simulation of replica 1 using the hbond module of pytraj[153]. Hydrogen bonds were defined by a donor-hydrogen-acceptor angle of at least 135° and a donor-acceptor distance less than 3 Å. The distances $d$ between donor and acceptor atom pairs were scaled according to $d_{scaled} = (d - 1.8\ \text{Å})^2$ and then used as input features to train a random forest classifier with the different classes corresponding to the eight possible isomerization states of BK. The scaling proved to be helpful for the subsequent clustering as lower distances get a higher weight and the density of higher-distance noise gets reduced. As before the scikit-learn software[191] was used. The distances $d_{scaled}$ were ranked by feature importance to find those that are best suited for discrimination between the eight isomerization states of BK and subsequently used as input for HDBSCAN clustering. This was done to find distinct conformations that may possibly be exclusively populated by a particular BK isomerization state. All found clusters

according to this metric were validated by root-mean-square-deviation (RMSD) analysis (Supporting Information (SI) Figs. B.1 - B.3 A). Note that we found the present approach of clustering based on important hydrogen-bond contacts to be more suited for clustering a partially unstructured peptide than a standard backbone-RMSD based approach. This is due to the large variety of backbone conformations accessible to aqueous BK, in particular concerning the random-coil like N-terminal region.

For additional analyses, the trajectory of replica 1 was filtered with respect to the eight different isomerization states of BK using dihedral angle ranges of 180±45° for the trans state and 0 ± 45° for the cis state. The resulting trajectories for states TTT, TTC, TCT, CTT, TCC, CTC, CCT and CCC contained 3236, 3687, 1845, 92, 3565, 6857, 1353 and 1642 frames, respectively and allowed independent investigation of the different isomerization states.

Molecular-Mechanics Generalized-Born Surface-Area (MMGBSA) free-energy calculations were performed on the different states using the MMPBSA.py software from the AmberTools software package [148]. The implicit water model igb = 5 was employed in combination with the mbondi radii set [192, 193]. A salt concentration of 0.1 M was used, the solvent relative dielectric permittivity was set to 80, the solute interior dielectric permittivity was set to 1 and nonpolar solvation free energy contributions were calculated [194] based on a surface tension coefficient of 0.0072 kcal·mol$^{-1}$·Å$^{-2}$. Entropy calculation was omitted due to the high computational demand and the introduction of large statistical uncertainties. Energies were decomposed on a per-residue basis for further insight [148].

The different states were also analyzed with respect to violations of the upper bounds of interproton distance ranges corresponding to measured Nuclear Overhauser Effect (NOE) intensities for zwitterionic BK in dimethylsulfoxide containing 1% water at 300 K [171]. The violation $\Delta_{ij}$ for a proton pair $i$, $j$ was calculated as the inverse $6^{th}$-power averaged proton-proton distance, minus the upper bound $d_{ij}^{NOE,exp}$ of the experimental NOE,

$$\Delta_{ij} = \langle r_{ij}^{-6} \rangle^{-1/6} - d_{ij}^{NOE,exp} \quad , \tag{4.14}$$

where $r_{ij}$ are instantaneous distances between protons $i$ and $j$, and the angular brackets denote ensemble averaging over the trajectories pertaining to the different isomerization states. The upper bounds were set to $d_{ij}^{NOE,exp} = 3$, 3.5, 4 or 4.5 Å for strong, medium, weak or very weak NOEs, respectively. Visualization of structures and trajectories was performed using VMD [154]. Throughout, errors

on free energies were calculated as the standard error of the mean of values from samples obtained by dividing the data in 5 equally-sized subsets.

## 4.3 Results and Discussion

### 4.3.1 Enhanced Sampling of Cis/Trans Isomerization by $\omega$BP-REMD

The $\omega$BP-REMD scheme was capable of inducing frequent cis/trans transitions in all the Pro residues of BK. Enhanced isomerization rates in the upper replicas propagated down the replica ladder. Thus, excellent cis/trans transition rates of 11.3, 10.6 and 9.6 $ns^{-1}$ for Pro2, Pro3 and Pro7, respectively, were observed in replica 1 (Fig. 4.2 A, top). Enhanced sampling of isomerization events is important because it allows a statistically meaningful analysis of the cis/trans equilibrium via Boltzmann inversion of corresponding probabilities (Eqs. 4.3 and 4.7; Section 4.2.2). The barrier height $V_{\omega,1} = 1$ kcal·$mol^{-1}$ of the added cosine term with multiplicity $n = 1$ (Eq. 4.2; Section 4.2.1) is suitable to achieve balanced cis/trans sampling in all three Pro residues. This is evidenced by the measure of the running mean of isomerization states, (Eq. 4.13; Section 4.2.4), where trans states are mapped to "1" and cis states are mapped to "0". The running mean of isomerization states indicates a small equilibration time before stable fluctuations around 0.5 without slow drifts towards any isomerization state during more than 800 ns of simulation time occur (Fig. 4.2 A, bottom).

Even if the sampling of cis/trans isomerization of any Pro residue in BK is drastically enhanced compared to an unbiased simulation setup, the fact that three Pro residues occur warrants further diagnostics to analyze the extent of sampling. For example, it could be that cis/trans isomerization events of the Pro residues are heavily correlated meaning that e.g. only states TTT and CCC of all the eight possible isomerization states of BK would be accessed if all Pro residues always switched simultaneously between the isomerization states from trans to cis and back. Here, however, isomerizations of the different Pro residues occur with only marginal linear correlations as shown by the very low Pearson correlation coefficients, $R(\mathrm{Pro2}, \mathrm{Pro3})$ = -0.14, $R(\mathrm{Pro2}, \mathrm{Pro7})$ = 0.29 and $R(\mathrm{Pro3}, \mathrm{Pro7})$ = -0.15.

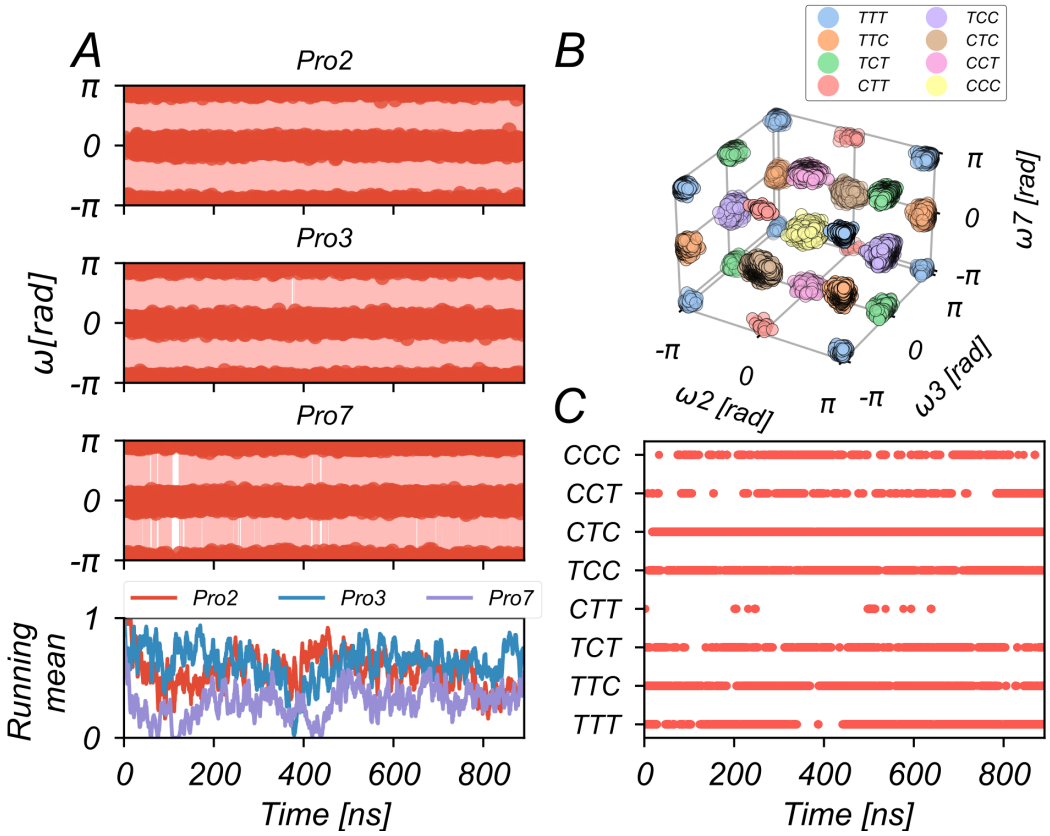



**Fig. 4.2.** Cis/trans isomerization of the Pro residues (Pro2, Pro3 and Pro7) in BK investigated with the $\omega$BP-REMD method. (A) Time series of the three $\omega$ dihedral angles sampled in replica 1 during a $\omega$BP-REMD simulation of 885 ns length (top). The data is also depicted as a running mean of sampled states (Eq. 4.13), where trans states have been assigned a "1" and cis states a "0" (bottom). (B) Illustration of the isomerization states, sampled in replica 1 in the three-dimensional space given by values of the $\omega$ dihedral angle of Pro2, Pro3 and Pro7 ($\omega_2$, $\omega_3$ and $\omega_7$, respectively). Each dot represents a sampled simulation frame. The isomerization states of BK are labelled TTT, TTC, TCT, CTT, TCC, CTC, CCT and CCC, where letters "T" and "C" refer to the trans and cis isomerization states, respectively of single Pro residues, and the first, second and third positions refer to Pro2, Pro3 and Pro7, respectively. (C) Time series of the isomerization states of BK sampled in replica 1.

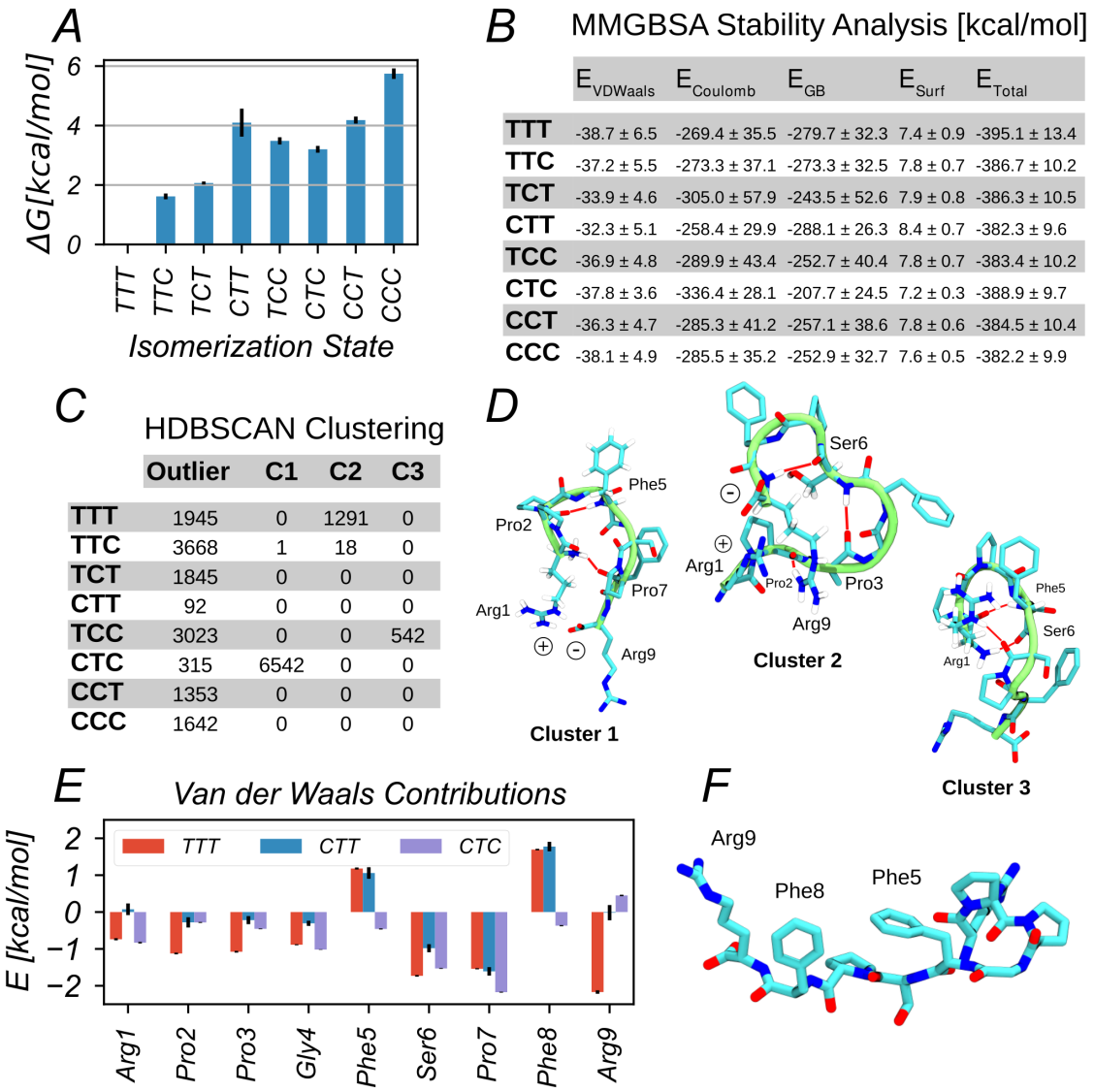
Thus all of the eight possible isomerization states of BK have been visited in multiple back and forth transitions as shown in Figs. 4.2 B and C. The percentages of simulation time in replica 1 spent in the CCC, CCT, CTC, TCC, CTT, TCT, TTC and TTT states are 7.4, 6.1, 30.8, 16.0, 0.4, 8.2, 16.6 and 14.5%, respectively. Although state CTT is accessed rather infrequently, the fact that its occurrences are distributed throughout the total simulation length of 885 ns supports simulation convergence and allows meaningful extraction of free energies.

In summary, $\omega$BP-REMD allows exhaustive sampling of the important phase-space regions in replica 1 for the investigation of cis/trans equilibria in BK, even in the multidimensional isomerization process with three degrees of freedom, the peptide bond dihedral angles of Pro2, Pro3 and Pro7. Configurations that would be very unfavorable in unbiased simulations, were sampled extensively.

### 4.3.2 Free Energy of BK Isomerization States

The unbiased probability densities of sampling in replica 1 (Eq. 4.3; Section 4.2.2) suggest that zwitterionic BK in water mainly adopts the all-trans configuration with a probability of $P(\omega_2 = \omega_3 = \omega_7 = 180^\circ)$ =89.99 $\pm$ 0.98%. The corresponding free energy is obtained from Boltzmann inversion (Eq. 4.7). It was here set to zero to express the free energies of the other isomerization states relative to the all-trans state. The resulting relative free energies for all states are listed in Fig. 4.3 A. These states are more unfavorable than the all-trans state by 1.62-5.75 kcal·mol$^{-1}$, which is about 3-10 times the value of $k_BT$ at 300 K. The all-trans state also populates configurations with the most favorable average interaction energies of $E$ = -395.1 $\pm$ 13.4 kcal·mol$^{-1}$ as shown by MMGBSA stability analysis (Fig. 4.3 B). Most notably, it is favorable solvent interactions which promote the all-trans state. Electrostatic BK-solvent interactions in the all-trans state are more favorable by 2-26% than in the other states, except CTT which presents even more favorable solvent interactions (Fig. 4.3 B).

HDBSCAN clustering of important hydrogen-bond contacts identifies a cluster of similar configurations (cluster 2 in Figs. 4.3 C and D) in the highly diverse set of BK configurations, populated primarily by all-trans states and some TTC states. The latter is the state closest in free energy to the all-trans state (1.62 kcal·mol$^{-1}$ more unfavorable; Fig. 4.3 A) and has a similarly favorable MMGBSA




| | $E_{VDWaals}$ | $E_{Coulomb}$ | $E_{GB}$ | $E_{Surf}$ | $E_{Total}$ |
|---|---|---|---|---|---|
| TTT | -38.7 ± 6.5 | -269.4 ± 35.5 | -279.7 ± 32.3 | 7.4 ± 0.9 | -395.1 ± 13.4 |
| TTC | -37.2 ± 5.5 | -273.3 ± 37.1 | -273.3 ± 32.5 | 7.8 ± 0.7 | -386.7 ± 10.2 |
| TCT | -33.9 ± 4.6 | -305.0 ± 57.9 | -243.5 ± 52.6 | 7.9 ± 0.8 | -386.3 ± 10.5 |
| CTT | -32.3 ± 5.1 | -258.4 ± 29.9 | -288.1 ± 26.3 | 8.4 ± 0.7 | -382.3 ± 9.6 |
| TCC | -36.9 ± 4.8 | -289.9 ± 43.4 | -252.7 ± 40.4 | 7.8 ± 0.7 | -383.4 ± 10.2 |
| CTC | -37.8 ± 3.6 | -336.4 ± 28.1 | -207.7 ± 24.5 | 7.2 ± 0.3 | -388.9 ± 9.7 |
| CCT | -36.3 ± 4.7 | -285.3 ± 41.2 | -257.1 ± 38.6 | 7.8 ± 0.6 | -384.5 ± 10.4 |
| CCC | -38.1 ± 4.9 | -285.5 ± 35.2 | -252.9 ± 32.7 | 7.6 ± 0.5 | -382.2 ± 9.9 |

| | Outlier | C1 | C2 | C3 |
|---|---|---|---|---|
| TTT | 1945 | 0 | 1291 | 0 |
| TTC | 3668 | 1 | 18 | 0 |
| TCT | 1845 | 0 | 0 | 0 |
| CTT | 92 | 0 | 0 | 0 |
| TCC | 3023 | 0 | 0 | 542 |
| CTC | 315 | 6542 | 0 | 0 |
| CCT | 1353 | 0 | 0 | 0 |
| CCC | 1642 | 0 | 0 | 0 |

**Fig. 4.3.** Free energies and representative conformations for the different isomerization states of BK. The isomerization states of BK are labelled TTT, TTC, TCT, CTT, TCC, CTC, CCT and CCC, where letters "T" and "C" refer to the trans and cis isomerization states, respectively of single Pro residues, and the first, second and third positions refer to Pro2, Pro3 and Pro7, respectively. (A) Free energies of the isomerization states relative to a value of zero for the all-trans state. The underlying values $G(state)$ were obtained according to Eq. 4.7 (Section 4.2.2). Error bars refer to the standard error of the mean of subset free-energy values. (B) MMGBSA stability analysis of isomerization states. $E_{\mathrm{VDWaals}}$, $E_{\mathrm{Coulomb}}$, $E_{\mathrm{GB}}$ and $E_{\mathrm{Surf}}$ refer to the average molecular mechanics van-der-Waals and Coulomb energies as well as Generalized-Born electrostatic solvation free energy and nonpolar surface area-based solvation free energy, respectively. Differences in other energy terms like $E_{\mathrm{Bond}}$, $E_{\mathrm{Angle}}$ and $E_{\mathrm{Dihedral}}$ are insignificant between the states and these energy terms are hence omitted from the table for better clarity. $E_{\mathrm{Total}}$ is the sum of all energy terms. Errors refer to the standard error of the mean of subset quantities.

**Fig. 4.3.** Continuation of figure caption. (C) HDBSCAN clustering on the combined isomerization states sampled in replica 1. For each isomerization state, the number of members in the three most occupied clusters (C1, C2 and C3) is reported, along with the number of structures not present in any of those ("outliers"). (D) Representative configurations for clusters C1, C2 and C3 shown in stick representation. Important salt bridges and hydrogen bonds and the involved residues are highlighted. Backbone alignment is illustrated by green-tube representation. (E) Residue-wise decomposition of the MMGBSA van-der-Waals energy contribution $E_{\mathrm{VDWaals}}$ reported in panel (B) for selected isomerization states. Error bars refer to the standard error of the mean of subset energies. (F) Exemplary conformation of state CTT with unfavorable van-der-Waals energy, shown in stick representation. Important residues are highlighted.

electrostatic solvation contribution (2% more unfavorable; Fig. 4.3 B). In the all-trans state, configurations fall into cluster 2 with a probability of $\sim$ 40%. Backbone root-mean-square deviations to cluster 2 are shown in SI Fig. B.2 A.

Structures of cluster 2 are characterized by a salt bridge of the C-terminal carboxylate group to the positively charged guanidinium group of Arg1 as well as hydrogen-bond interactions between the backbone oxygen atom of Pro2 and the side chain guanidinium group of Arg9 (Fig. 4.3 D). Furthermore, frequent hydrogen-bond contacts between the backbone oxygen atom of Pro3 and the backbone amide hydrogen atom of Ser6, as well as between the backbone oxygen atom of Ser6 and the amide hydrogen atom of Arg9 are established in cluster 2 (Fig. 4.3 D). The latter hydrogen bond is present in 30.8% of configurations sampled in the all-trans state, while it is only present in 7.2% of simulation frames for state TCT, 4.3% for CTT and 10.7% for CCT (using a looser donor-acceptor distance criterion of 3.5 Å; Section 4.2.4 and SI Table B.1) and it is absent in the remaining isomerization states of BK. It is important in establishing the $\beta$-turn structure in the C-terminal residues Ser6-Arg9 of BK that was also found in in-solution NMR experiments of BK (Fig. 4.3 D and SI Fig. B.2 B) [167, 169–171].

Our finding of a dominant all-trans population is supported by numerous NMR experiments,[170, 171, 183, 195–197] which allow determination of a Pro isomerization state via the chemical shifts of Pro C$\beta$ and C$\delta$ atoms [198].

Furthermore, we analyzed the conformations sampled in replica 1 with respect to NOE upper bound violations using experimental NOE intensities provided in Ref. [171] for zwitterionic BK in solution. This provides experimental validation for the conformations sampled in the all-trans state of BK in the present study (Table 4.1). In particular, for a total of 30 NOE intensities, no violations were found for the conformations pertaining to the all-trans state. Also, no violations were found for state CTT. Note, in this context, that none of the experimental NOEs involved

protons of Pro2. However, NOEs involving both Pro3 and Pro7 were violated for states CCC, TCC and TCT, NOEs involving Pro3 only were violated for state CCT, and NOEs involving Pro7 only were violated for states CTC and TTC (Table 4.1). This may be an additional indication that the experimental ensemble of structures presents Pro3 and Pro7 in the trans states.

| state | atoms | | NOE intensity | $\Delta_{ij}$ [Å] |
|---|---|---|---|---|
| CCC | Phe5, H | Pro3, HA | vw | 0.6 |
| | Arg9, H | Pro7, HA | vw | 0.1 |
| CCT | Phe5, H | Pro3, HA | vw | 0.7 |
| CTC | Phe8, H | Pro7, HD2[a] | w | 1.4 |
| TCC | Phe5, H | Pro3, HA | vw | 0.6 |
| | Arg9, H | Pro7, HA | vw | 0.2 |
| TCT | Phe5, H | Pro3, HA | vw | 0.8 |
| | Arg9, H | Pro7, HA | vw | 0.2 |
| TTC | Arg9, H | Pro7, HA | vw | 0.3 |

**Tab. 4.1.** NOE upper bound violations $\Delta_{ij}$ (Eq. 4.14) for different isomerization states of BK. The isomerization states of BK are labelled TTT, TTC, TCT, CTT, TCC, CTC, CCT and CCC, where letters "T" and "C" refer to the trans and cis isomerization states, respectively of single Pro residues, and the first, second and third positions refer to Pro2, Pro3 and Pro7, respectively. No violations are observed for states CTT and TTT. NOE intensities are given as strong (s), medium (m), weak (w), or very weak (vw) and corresponding experimental upper bounds [171] were 3, 3.5, 4, or 4.5 Å, respectively.

Besides the all-trans state, the TTC and TCT states are populated with unbiased probabilities of 6.18 $\pm$ 0.81% and 2.82 $\pm$ 0.21%, respectively. These states are closest in free energy to the all-trans state (Fig. 4.3 A). MMGBSA stability analysis does not reflect a clear energetic preference of TTC compared to other non-all-trans states, except for the electrostatic solvation free-energy contribution that is of similarly low magnitude as that for the all-trans state (Fig. 4.3 B). Favorable solvation or entropic contributions omitted in the MMGBSA analysis may hence play a role in stabilizing the TTC state. Note that only a tiny fraction of TTC configurations (0.5%) could be assigned to certain conformational clusters (Fig. 4.3 C).

The CTT state is significantly less probable, with a probability of 0.19 $\pm$ 0.09%. While it displays the most favorable MMGBSA electrostatic solvation free energy it has particularly unfavorable intrapeptide nonpolar interactions (Fig. 4.3 B). Notably, the cis isomer of Pro2, as it occurs in state CTT, appears to be significantly more unfavorable than the cis isomer of Pro3 or Pro7 (e.g. as occurring in states TCT and TTC). This may be traced back to unfavorable van-der-Waals interactions involving

Phe5 and Phe8 (Fig. 4.3 F). As indicated by the residue-wise decomposition of MMGBSA energy contributions, these residues have particularly high van-der-Waals energy contributions in states TTT and CTT, whereas their van-der-Waals energy contributions are favorable in state CTC (Fig. 4.3 E). This is also reflected in the average van-der-Waals interaction energy between the two residues, which is, among all the BK isomer states, the most favorable in state CTC (-2.0kcal·mol$^{-1}$; SI Table B.2).

The unfavorable intrapeptide energies observed in state CTT can be largely ameliorated by isomerization of a second Pro residue, Pro7. In fact, the CTC state of BK has the third lowest relative free energy with respect to the all-trans state and the second lowest MMGBSA total energy besides the all-trans state. The stability of state CTC is particularly reflected in a low average Coulomb energy of -336.4 $\pm$ 28.1 kcal·mol$^{-1}$ at expense of slightly more unfavorable solvent interactions (Fig. 4.3 B).

Conformational clustering shows that state CTC is highly structured with almost all ($\sim$ 95%) sampled configurations falling into cluster 1. This cluster appears structurally very homogeneous (SI Fig. B.1 B). Backbone root-mean-square deviations to cluster 1 are shown in SI Fig. B.1 A. In cluster 1, a very stable salt bridge is formed between the C-terminal carboxylate group and the positively charged guanidinium group of Arg1 (Fig. 4.3 D). This is in keeping with the low Coulomb energy of state CTC (Fig. 4.3 B). However, as suggested by the structural homogeneity of cluster 1 (SI Fig. 1 B), unfavorable conformational entropy contributions may act to destabilize state CTC. Hydrogen bonding between the backbone oxygen atom of Pro2 and the backbone amide hydrogen atom of Phe5 as well as between the backbone oxygen atom of Pro7 and the positively-charged N-terminal amino group further stabilizes the conformation.

A smaller cluster 3 populated by conformations of state TCC was found. Cluster 3 appears to be structurally very homogeneous (SI Fig. B.3 B). It features frequent hydrogen bonding between the backbone oxygen atom of Arg1 and the backbone amide hydrogen atom of Phe5, as well as between the backbone oxygen atom of Ser6 and the positively-charged N-terminal amino group and the backbone oxygen atom of Phe5 and the guanidinium group of Arg1. Backbone root-mean-square deviations to cluster 3 are shown in SI Fig. B.3 A.

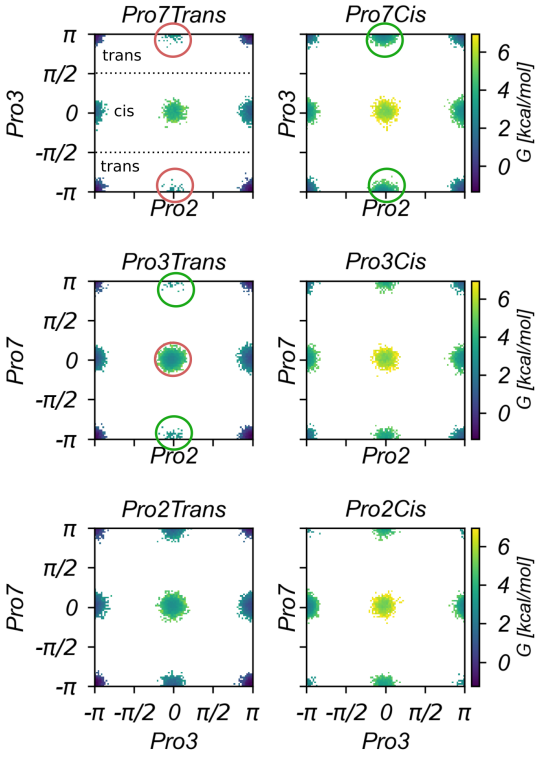
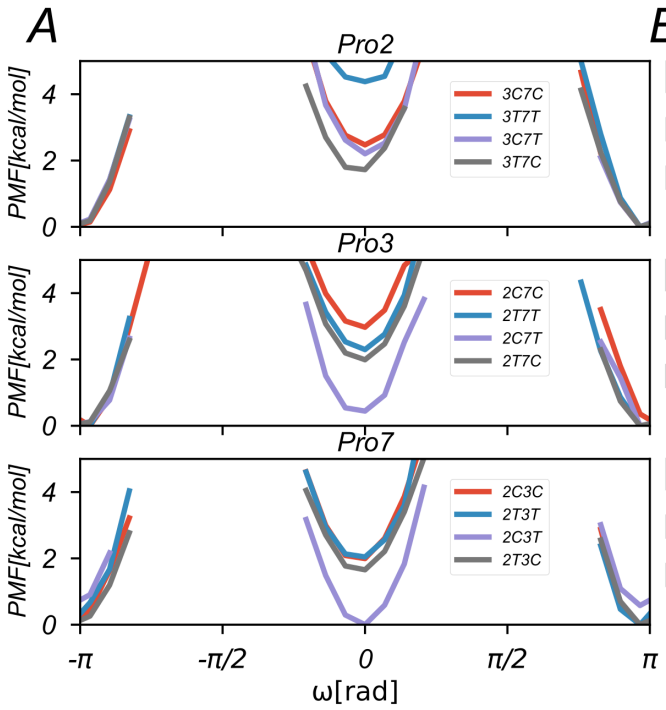



**Fig. 4.4.** Two-dimensional free-energy landscapes as function of the $\omega$ dihedral angles of two Pro residues $i$ and $j$, indicated on the $x$- and $y$-axes, of BK. The free energies were obtained from Boltzmann inversion of the corresponding unbiased probabilities (Eq. 4.8). They depend on the isomerization state of the third Pro residue $k$, which is indicated on top of each graph. The highlighted regions show that Pro2 is significantly more likely to be found in the cis isomer, if Pro7 also adopts the cis isomer.

### 4.3.3 Cis-Trans Isomerization Equilibria of all Pro Residues in BK

Based on a two-dimensional projection of the three-dimensional unbiased probability densities (Eq. 4.8; Section 4.2.2), two-dimensional free-energy surfaces were calculated via Boltzmann inversion (Fig. 4.4). This allows visual detection of the coupling of isomerization events. For instance, Pro2 appears to be significantly more likely to be found in the cis isomer, if Pro7 is also in the cis isomer. This is consistent with the above noted observation that BK conformations achieved by isomerization of Pro7 to the cis state (as in state CTC) appear to be much more favorable than those compatible with only Pro2 being in the cis state (as in state CTT; Section 4.3.2).




B

Pro2 Isomerization

| State | $\Delta G_{C \to T}$ [kcal/mol] | $\Delta\Delta G$ |
|---|---|---|
| **3Trans7Trans** | -4.09 ± 0.47 | 0 |
| **3Trans7Cis** | -1.58 ± 0.15 | 2.51 ± 0.49 |
| **3Cis7Trans** | -2.11 ± 0.13 | 1.98 ± 0.48 |
| **3Cis7Cis** | -2.26 ± 0.21 | 1.83 ± 0.51 |

Pro3 Isomerization

| State | $\Delta G_{C \to T}$ [kcal/mol] | $\Delta\Delta G$ |
|---|---|---|
| **2Trans7Trans** | -2.06 ± 0.05 | 0 |
| **2Trans7Cis** | -1.86 ± 0.16 | 0.20 ± 0.16 |
| **2Cis7Trans** | -0.09 ± 0.48 | 1.97 ± 0.48 |
| **2Cis7Cis** | -2.54 ± 0.21 | -0.48 ± 0.22 |

Pro7 Isomerization

| State | $\Delta G_{C \to T}$ [kcal/mol] | $\Delta\Delta G$ |
|---|---|---|
| **2Trans3Trans** | -1.61 ± 0.09 | 0 |
| **2Trans3Cis** | -1.41 ± 0.13 | 0.20 ± 0.16 |
| **2Cis3Trans** | 0.89 ± 0.48 | 2.50 ± 0.49 |
| **2Cis3Cis** | -1.56 ± 0.21 | 0.05 ± 0.23 |

**Fig. 4.5.** Conditional free-energy profiles and isomerization free energies for BK. (A) Free-energy profiles as function of the $\omega$ dihedral angle of a Pro residue $i$, indicated on top of each graph. The free energies were obtained from Boltzmann inversion of the corresponding unbiased probabilities (Eq. 4.9). They depend on the isomerization state of the other two Pro residues $j$ and $k$, which are indicated as labels to each PMF, where numbers 2, 3 and 7 denote the Pro residue and letters "T" and "C" denote the trans or cis state for the Pro residue. (B) Conditional isomerization free energies $\Delta G_{\mathrm{cis}\to\mathrm{trans}}$ (Eq. 4.12) for Pro residues 2, 3 and 7 (from top to bottom). The isomerization state of the other two Pro residues is indicated in bold font. The quantity $\Delta\Delta G$ describes how the conditional isomerization free energy changes when any of the other two Pro residues switches from the trans to the cis state. Errors refer to the standard error of the mean of subset quantities.

By calculating PMFs along the $\omega$ dihedral angles of each Pro residue of BK in dependence of the other Pro isomerization states, we can systematically and in detail quantify the mutual influence of Pro isomerization states on the isomerization equilibria in BK via conditional isomerization free energies (Eq. 4.12; Section 4.2.3).
These show that isomerization processes in BK are not independent (Fig. 4.5 A).

For instance, for Pro2, 4.09 kcal·mol$^{-1}$ have to be invested when switching from the trans to cis state if both Pro3 and Pro7 are in the trans state, but only 1.58 kcal·mol$^{-1}$ have to be invested when Pro3 is in the trans state and Pro7 is in the cis state. Interestingly, when Pro3 is in the cis state, the isomerization of Pro7 has almost no influence on the isomerization free energy of Pro2 (Fig. 4.5 B, top).

Pro3 shows the least discrimination between the cis and trans states when Pro2 is in the cis state and Pro7 is in the trans state. In this case, the difference between the Pro3 cis and trans states is only -0.09 kcal·mol$^{-1}$. Otherwise, switching from the cis to the trans state involves a free energy gain of -2.54, -2.06 and -1.86 kcal·mol$^{-1}$ for states where both Pro2 and Pro7 are in the cis state, both are in the trans state, or Pro2 is in the trans and Pro7 in the cis state, respectively (Fig. 4.5 B, middle).

Of note, Pro7 is more likely to be found in the cis isomer when Pro2 is in the cis isomer with a free energy of isomerization of 0.89 kcal·mol$^{-1}$ for switching to the trans isomer. This appears to be the only conditional isomerization equilibrium of zwitterionic BK in water which favors the cis state of a Pro residue. In contrast, other cis/trans combinations of Pro2 and Pro3 entail a more favorable trans state of Pro7, with free energy gains of -1.5 kcal·mol$^{-1}$ (Fig. 4.5 B, bottom).

All Pro residues $i$ considered, the differences in conditional isomerization free energies with respect to Pro residues $j$ and $k$ being in the trans state span a range of -0.48 to 2.51 kcal·mol$^{-1}$ (Fig. 4.5 B). Isomerization equilibria can hence be drastically influenced by the isomerization states of other Pro residues. To our knowledge, this is the first time that a strong coupling of multiple Pro isomerization states has been systematically illustrated and quantified via free-energy calculations. The fact that such coupling is observed in BK, a short and highly flexible peptide with partial random-coil-like character is remarkable. Especially since two of its Pro residues (Pro2, Pro3) are located in the unstructured N-terminal region (Arg1-Phe5), one may assume that coupling of Pro isomerization states may play a similar, if not more decisive role in systems where conformational restraints are present, as e.g. in folded proteins. For instance, stable secondary structure elements can be adversely affected by the isomerization of a given Pro residue and their destabilization may facilitate the isomerization of another Pro residue. On the

other hand, isomerization of a given Pro residue to the native state, as in a folding process, may stabilize certain secondary structure elements and concomitantly also favor that isomerization state of another Pro residue which is more compatible with the present structural restraints.

## 4.4 Conclusion

The cis/trans isomerization of Pro can significantly impact the protein folding process [109] and contribute to complex biological phenomena, such as molecular timing [119]. This is attributed to the high energy barrier (approximately 20 kcal·mol$^{-1}$ ) and the consequential extended timescales (seconds to minutes) associated with Pro cis/trans isomerization in nature. Despite the biological relevance of these timescales, they are not accessible in atomistic explicit-solvent MD simulations. Therefore, we employed the $\omega$BP-REMD method in this study to investigate Pro cis/trans isomerization.

In Chapter 3, we introduced $\omega$BP-REMD as an enhanced-sampling method, allowing efficient sampling of cis/trans transitions of Pro residues in a manner that achieves approximately equal occupations of the cis and trans isomers. Isomerization free energies are calculated via Boltzmann inversion of unbiased probabilities. Besides its computational efficiency arising from equally focusing the sampling in the lowermost replica on the cis and trans isomerization states, an advantage of $\omega$BP-REMD is its power in the investigation of systems where multiple Pro residues occur. Unlike methods like metadynamics, where one has to sample the volume of the multi-dimensional conformational space defined by combining various reaction coordinates, $\omega$BP-REMD independently biases each of these multiple reaction coordinates. Hence, when several reaction coordinates are considered simultaneously, $\omega$BP-REMD requires considerably less sampling effort compared to standard metadynamics or Umbrella sampling approaches.

The system studied in this work was zwitterionic aqueous BK, a partially unstructured peptide with three Pro residues. The $\omega$BP-REMD simulation results, biased probabilities of BK isomerization states, were unbiased and analyzed through three-, two-, and one-dimensional free-energy surfaces. Conditional isomerization free energies, depicting the influence of the isomerization state of two other Pro residues on a specific Pro residue, were calculated for all Pro residues in BK. Thus, comprehensive quantitative insight into the coupling of isomerization states can be

obtained. Furthermore, the sampled conformations were analyzed in detail and compared to experimental NMR data.

$\omega$BP-REMD not only provided excellent statistics for cis/trans transitions in all three BK Pro residues but also enabled the exploration of all eight possible BK isomerization states with converged probabilities over a simulation period of 885 ns. Key findings are:

1. The all-trans state is found to be the preferred state. The preference for the all-trans state of aqueous zwitterionic BK has been validated by several experimental NMR studies. In present study, an analysis of NOE upper bound violations further supported the conformational sampling of the all-trans state.

2. Distinct conformational clusters are sampled by the eight different isomerization states of BK. Of note, in about one third of the simulation frames, the all-trans state presents the characteristic Ser6-Arg9 hydrogen bond forming the experimentally-confirmed C-terminal $\beta$-turn motif. The cluster predominatly sampled by state CTC appears highly structured. While state CTC is about 3.3 kcal·mol$^{-1}$ less favorable than the all-trans state, it involves profound stabilization by favorable intrapeptide electrostatic interactions.

3. Isomerization of the three Pro residues in BK can occur in a coupled fashion, i.e. the isomerization state of a given Pro residue can influence the cis/trans equilibrium of other Pro residues. We quantify the influence of the isomerization state of two Pro residues $j$ and $k$ on the isomerization equilibrium of Pro residue $i$ with conditional isomerization free energies, which are the difference in the free energies of the cis and trans isomerization state of Pro $i$ for the different combinations of isomerization states of Pro residues $j$ and $k$. These free energies are conveniently obtained from projection of the three-dimensional (unbiased) probability density obtained from replica 1 of a $\omega$BP-REMD simulation on the reaction coordinate of interest, i.e. the $\omega$ dihedral angle of Pro $i$. A notable example for BK is that Pro7 is more likely to be found in the cis isomer than in the trans isomer when Pro2 is in the cis isomer. This appears to be the only conditional isomerization equilibrium of zwitterionic BK in water which favors the cis state of a Pro residue.

4. The influence of Pro isomerization states on the cis/trans equilibrium of other Pro residues may be large, i.e. conditional isomerization free energies may differ by up to 2.5 kcal·mol$^{-1}$.

While it was previously noted that the isomerization states of the individual Pro residues in BK appear not to be independent [164], to our knowledge, the present work is the first time that a strong coupling of Pro isomerization states has been systematically illustrated and quantified via extensive free-energy calculations. It is noteworthy that such coupling is observed in as small and unstructured a system as BK. This suggests that coupling of Pro isomerization states may play a similar, if not more significant role in systems exhibiting more conformational restraints like folded proteins.

# 5

# Ligand Binding and Global Adaptation of the GlnPQ Substrate Binding Domain 2 [1]

Substrate-binding domains (SBD) are crucial structural components of substrate transporters, mediating the transport of essential molecules across the cell membrane. The SBD2 domain of the glutamine (GLN) transporter from bacteria comprises two segments, D1 and D2, that bind GLN in the space between the domains in a closed conformation.

In the absence of a ligand, SBD2 adopts an open conformation, with increased distance between its domains. Molecular dynamics (MD) simulations conducted without ligands did not reveal any closure of the open conformation within the MD time scale. Upon addition of GLN, multiple reversible binding and unbinding events occurred at the binding site on the D1 domain. Notably, however, this binding did not induce domain closure, suggesting that substrate binding and global domain closure are not simultaneous events. When commencing from the GLN-bound closed crystal structure, the SBD2 structure maintained a closed state. However, upon the removal of the GLN ligand, it swiftly transitioned to the open state. Free energy simulations designed to induce opening to closing showcased a barrier for closure, both in the presence and absence of the ligand, with a significant penalty for closing when the GLN ligand was absent from the binding pocket.

Furthermore, simulations involving a Leu480Ala mutation revealed that the interaction between the C-terminal D1-tail$_{471-484}$ and a D2-helix$_{418-427}$, not directly contacting the substrate-binding region, played a pivotal role in governing the barrier of conformational switching in the SBD2 protein. Collectively, these results allow us to derive a model of the molecular mechanism of substrate binding to SBD2 and associated conformational changes.

[1] Apart from minor modifications, this chapter was originally published in [199].

## 5.1 Introduction

ATP-binding cassette (ABC) transporters constitute a vast superfamily of integral membrane proteins present across all kingdoms of life, playing a pivotal role in transporting essential molecules across the lipid cell membrane[4, 200–202]. These transporters are involved in crucial cellular processes, such as the uptake of amino acids and nutrients or the excretion of waste products into the external environment [203, 204]. ABC transporters are dimeric proteins that consist of two transmembrane domains (TMDs) and two cytoplasmic nucleotide binding domains (NBDs) as well as substrate-binding domains (SBDs). These SBDs mediate the initial binding of substrate and delivery to the translocation subunit [203, 205] and are often connected to the TMDs via a flexible linker [206–208].

Numerous three-dimensional (3D) structures of SBDs from various transporters have been determined in both unbound (apo) and substrate-bound (holo) states [209, 210]. Typically, SBDs consist of two domains (D1 and D2), with a ligand binding cavity in the space between them [211, 212]. In the absence of a ligand, an open domain arrangement is commonly adopted, while crystal structures in the presence of a substrate suggest a closed domain-domain conformation around the bound substrate [208, 210, 213]. Studies indicate that the transition to a closed state is facilitated by ligand binding [214–217].

Following substrate binding, the SBD complex binds to the translocator of the ABC transporter, initiating the translocation process to the cytoplasmic side of the membrane. One of the most extensively investigated ABC transporters is the GlnPQ transporter from bacteria that uses two SBDs (SBD1 and SBD2) for the import of asparagine, glutamine and glutamate: The SBD2 binds specifically to L-glutamine (GLN) and is essential for its subsequent transport [218].

High-resolution crystal structures of SBD2, both in the apo and holo forms with bound GLN, are available for the GlnPQ transporter [208] (see Fig. 5.1). Nuclear magnetic resonance (NMR) coupled with paramagnetic relaxation enhancement has been employed to characterize the open-close transition [219], suggesting a prevalence of the open state for the GLN binding protein in the absence of a ligand. In contrast, for the maltose binding protein, this technique implies a minor population of a closed form even in the absence of a ligand.

In addition, extensive single molecule Förster resonance transfer (smFRET) experiments have been performed on SBD2 in the presence and absence of GLN to elucidate the dynamics of the ligand binding and domain closing processes [204,

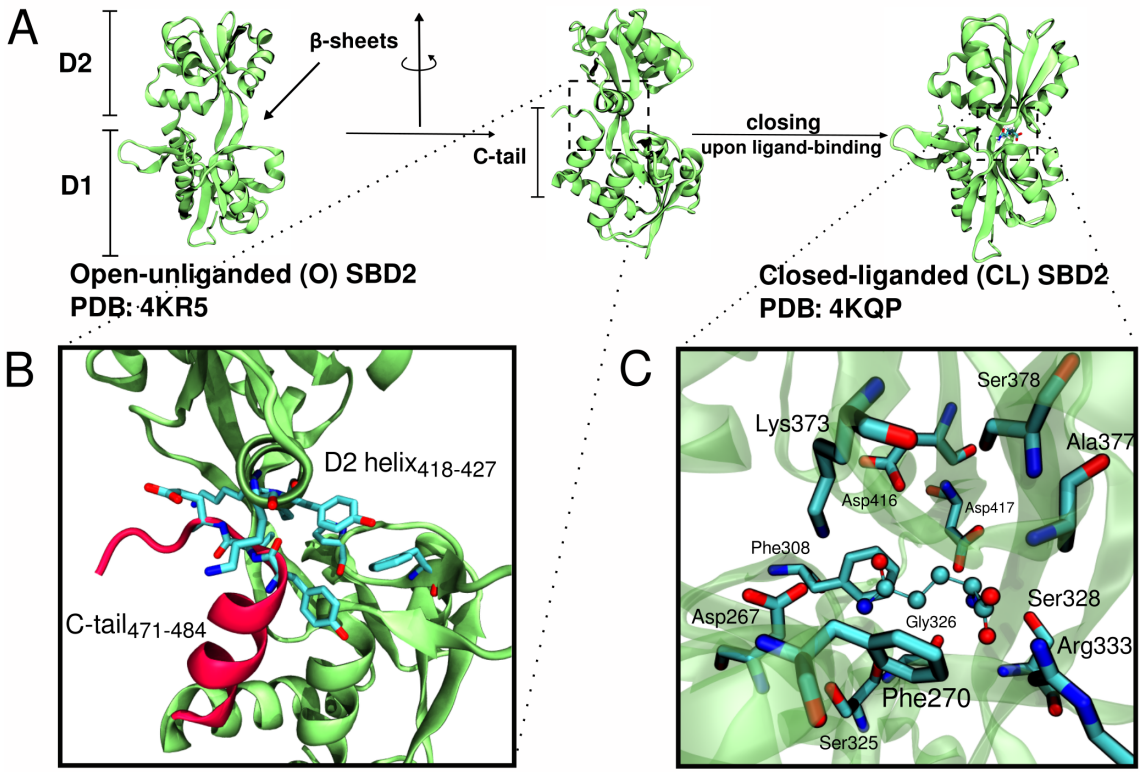



**Fig. 5.1.** (A) X-ray crystal structures of SBD2 of GlnPQ from Lactococcus lactis in open unliganded state and in the ligand(GLN)-bound closed conformation. The SBD2 is composed of two continuous subdomains D1 (residues 255-343, 444-484) and D2 (residues 349-438) connected via two anti-parallel $\beta$-sheets (residues 344-348, 439-443). (B) The C-terminal $\alpha$- helical region (C-tail: residues 471-484 in red, which are part of D1) interacts with D2 in the open-protein state and contacts residues in a D2 $\alpha$-helix (residues 418-427). (C) The second inset panel illustrates the crystal structure contacts of GLN in the bound closed complex.

217, 218, 220]. In the smFRET technique the global domain-domain motion can be detected reaching a time resolution in the millisecond regime and has allowed to elucidate the detailed kinetics of the ligand binding and associated global structural change. It has been shown that the overall transport cycle of GlnPQ is highly influenced by conformational dynamics of the SBDs [217]. However, it is neither possible to detect individual substrate binding events if these do not trigger a global change nor to characterize short lived intermediate states. Also, the detection of a smFRET signal may only allow one to distinguish different global structural changes qualitatively without further structural characterization on a molecular level [221, 222].

We employ Molecular Dynamics simulations on the SBD2 of the GlnPQ importer from the gram-positive bacterium Lactococcus lactis to investigate the ligand binding process and associated conformational changes in atomic detail. This technique has already been used to study the dynamics of GLN binding proteins [223]- [224] and related systems [225, 226] mostly in the apo state to investigate local and global mobility. Starting from the open unbound structure the addition of L-glutamine to the simulation box results in multiple reversible binding and unbinding events of the ligand to the larger of the two SBD2 domains (D1) forming basically the same contacts as in the known closed complex. Surprisingly, this does not result in domain closing on the time scale of the MD simulations (several $\mu$s). Starting from the closed form without ligand results, however, in a transition to the open form. Using extensive free energy simulations, we characterize the free energy profile for the open/close domain transition in the absence and presence of bound substrate and find both a stabilization of the closed form and lowering of the transition barrier in the presence of the ligand. The open form is stabilized by a helical segment located near the hinge region between the domains that forms non-covalent contacts not present in the closed arrangement. The in silico substitution of a critical Leucine at the contact interface results in a rapid closing in the presence of the substrate demonstrating that this interaction is key to the conformational switching behavior of the system. The simulations allow us to derive a model for the molecular mechanism of substrate binding and associated conformational changes.

## 5.2 Material and Methods

### 5.2.1 Unrestrained MD Simulations

The crystal structures of the open (PDB:4KR5) and the L-glutamine (GLN) bound closed SBD2 protein (PDB:4KQP) from Lactococcus lactis [208] served as start structures for the Molecular Dynamics (MD) simulations. The CUDA accelerated PMEMD [147] version of the AMBER16 software package [142] was used for all simulations employing the ff14SB force field parameters [141]. The protein molecules were solvated in explicit TIP3P water [143] in an octahedral box with a minimum distance of 10 Å to the box boundaries. The protonation state of the titratable amino acids were predicted via Poisson-Boltzmann calculations, using the free open-source web server H++ [149]. Simulations were started from the closed or open form in the absence or presence of GLN and a L480A mutation following the same equilibration protocol outlined below. In case of the MD simulations starting from the open apo SBD2 in the presence of substrate six L-glutamine molecules were placed randomly into the simulation box resulting in a ligand concentration of $\sim$ 25 mM. Sodium and chloride ions were added to neutralize the system and reach an ion concentration of 100 mM. After an energy minimization of 2000 steps of steepest descent, the systems were gradually heated up to 300 K while keeping positional restraints (force constant of 10 kcal·mol$^{-1}$ · Å$^{-2}$) on the protein non-hydrogen atoms during 1 ns simulation time. The restraints were gradually removed for another 1 ns followed by unrestraint production simulations of more than 3 $\mu s$. All bonds involving hydrogens were kept at optimal length using SHAKE [82]. The hydrogen mass reparation scheme was used allowing a time step of 4 fs [83]. Trajectory analysis was performed using the cpptraj and pytraj tools [153] of the AMBER16 suite. Visualization of structures and trajectories was performed using VMD [154].

### 5.2.2 Replica-Exchange Umbrella Sampling Simulations

Umbrella Sampling (US) simulations coupled with Hamiltonian Replica exchanges (H-REUS) between neighboring US intervals [89, 227] were performed to obtain the potentials of mean force for the opening/closing transition of SBD2 with and without L-glutamine present in the binding pocket. The distance between centers-of-mass of backbone atoms in two regions in the domains D1 and D2 of SBD2

served as a reaction coordinate $\xi$. The first center in the larger domain D1 contains the atoms CA, C, N of the residues 306-308, whereas the CA, C, N atoms of residues 396 and 397 form the other center-of-mass in D2. These segments are in close vicinity in the closed form (illustrated in Supporting Information, Figure C.6). A set of 16 US windows biased by harmonic potentials with force constants of 1 kcal·mol$^{-1}$· Å$^{-2}$ was generated with an equidistant spacing of 1 Åcovering distances from 6 Å up to 21 Å. During the replica exchange process exchanges between neighboring US windows were attempted every 1 ps reaching an acceptance rate of 25-45%. The equilibrium distance for the reaction coordinate in the unrestrained simulations of closed and open SBD2 structure was 7.7 Å and 20 Å, respectively. H-REUS simulations were carried out in the presence and absence of L-glutamine (GLN) in the binding pocket on the D1 domain. In each case one simulation was started from the open SBD2 configuration and the other starting from closed SBD2, resulting in a total of four H-REUS simulations. The overall simulation time for the ligand-bound state amounted to 15.1 $\mu s$ for all 16 US windows, where each US window was simulated for $\sim$ 400 ns for the first simulation and $\sim$ 540 ns for the second simulation, respectively. For the unbound state the total simulation time amounted to $\sim$ 16.6 $\mu s$, where each US window was simulated for $\sim$ 300ns for the first simulation and $\sim$ 741 ns for the second simulation. The simulated distributions along the reaction coordinates for the ligand-bound and unbound state were analyzed by employing the weighted histogram analysis method (WHAM) [103] yielding the corresponding free energy profile. Here, the implementation by Alan Grossfield [@228] was used which also allows error analysis via Monte Carlo Bootstrapping.

### 5.2.3 Evaluation of Trajectories Using the MMPBSA Technique

Average interaction energy calculations on the unbiased trajectories have been conducted using the MMPBSA trajectory post processing method [96], utilizing the MMPBSA.py program of the Amber package [97]. The purpose of the calculations was to obtain an estimate on the contribution of individual residues in the protein on the binding to the GLN ligand. Here, the 'single trajectory approach' was used, assuming that there are no significant conformational changes upon binding and yielding the mean interaction between substrate and receptor decomposed for each residue at the binding region. In each case 5000 frames in the ligand-bound state of

SBD2 sampled during free unrestrained MD simulation (Fig. 5.2) in the range from 1000ns up to 2500ns have been evaluated. The ion concentration was set to 100 mM (same as in the explicit solvent simulations). The harmonic conformational entropy contribution was included for calculating the total binding energy.

## 5.3 Results and Discussion

### 5.3.1 Simulation of the Open SBD2 in the Absence and Presence of L-glutamine

Unrestrained Molecular Dynamics (MD) simulations were started from the open SBD2 state (PDB: 4KR5). In the absence of L-glutamine (GLN) ligands the root-mean-square deviation (RMSD) with respect to the start structure stayed in the range of $\sim$ 2-3 Å within 1.8 $\mu s$ simulation time and no indication of spontaneous global domain closing. The RMSD with respect to the closed structure remained at a high level of $\sim$ 9-12 Å (see SI, Figure C.1). In a second simulation starting from the same SBD2 structure the simulation box contained six randomly placed L-glutamine (GLN) ligands. During the 3.3 $\mu s$ simulation the GLN sampled extensively the surface of the open protein (Figure 5.2 A, B). The GLN molecules stayed only a few nanoseconds at various surface positions (indicated by the large and rapidly fluctuating RMSD of most of the GLN molecules, Figure 5.2 B). However, several binding events to the binding pocket very close to the geometry found in the crystal structure of the closed ligand bound form of SBD2 (PDB: 4KQP) were captured. Remarkably, binding to the open form occurred only to the binding region on the large subdomain (D1: residues 255-343 and 444- 484) not contacting residues of the D2 (residues 349-438) domain. The lifetime of these complexes was much longer (ranging from 100-1000 ns) than for the non-specific placements distributed over the whole surface of the protein. After dissociation of one GLN, replacement with another GLN was observed two times indicating independent and uncorrelated binding events. All of the three captured complex formations resemble the ligand position in the crystallized placement and structure with an $\text{RMSD}_{lig} < 2.5$ Å (the RMSD of the ligand heavy atoms after best superposition of the receptor on the experimental reference structure, Figure 5.2 B, D). Surprisingly, on the time scale of the MD simulations no domain closing transition was observed. Very similar to the simulation in the absence of a ligand the RMSD of the sampled states remain

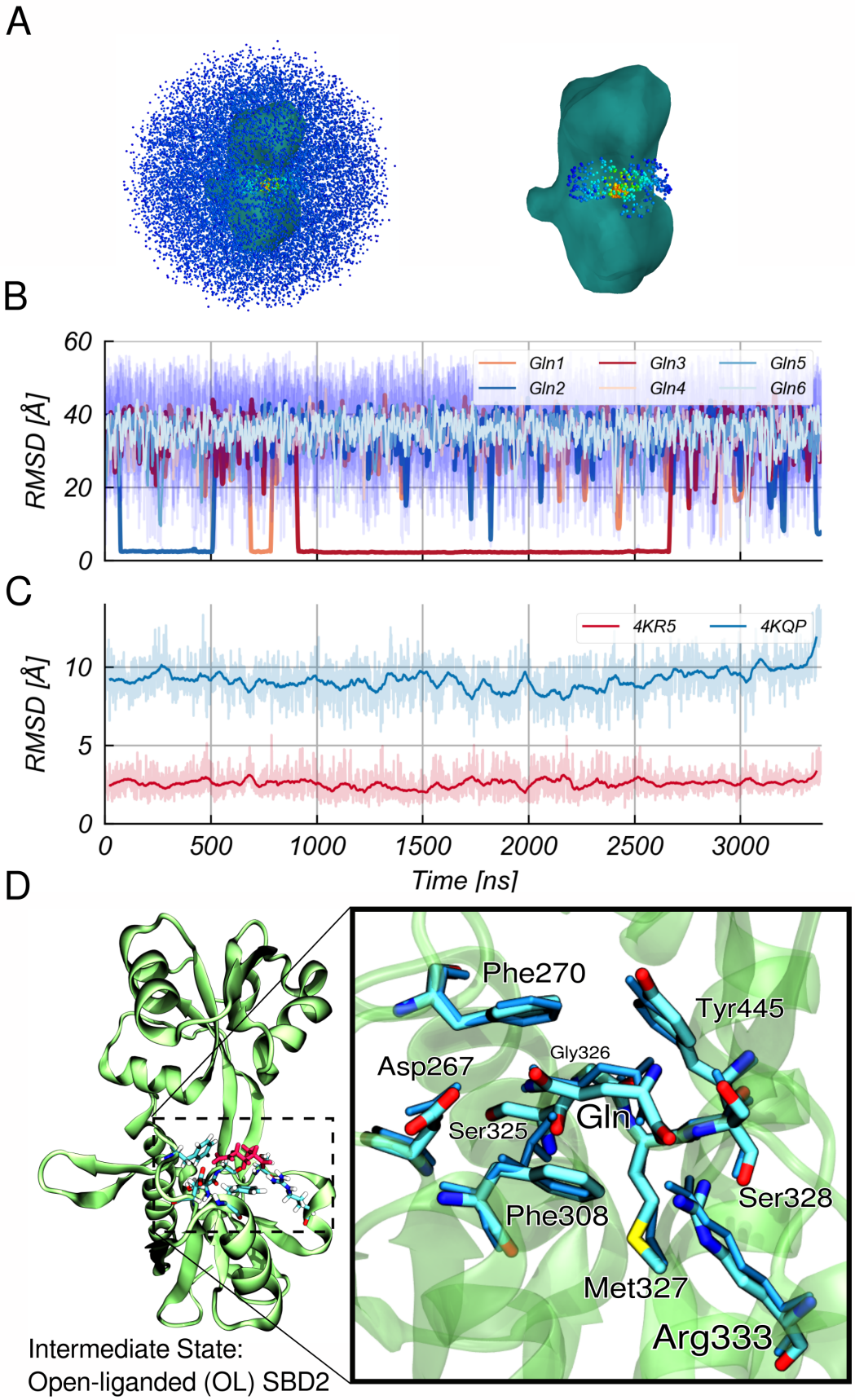



**Fig. 5.2.** Ligand-binding mechanism to the open-protein form of SBD2. (A) The blue dots (left panel) around the SBD2 structure (shown as surface contour) indicate positions sampled by the GLN ligand during the simulation starting with the open SBD2 and six randomly positioned GLN molecules. (Caption continued on next page.)

**Fig. 5.2.** In the right panel only the sampling near the binding site on the larger subdomain D1 is illustrated (the color of each dot indicates the sampling density, with red representing a high density and blue low density).(B) Root-mean-square deviation (RMSD, original data and running mean) of each GLN ligand (indicated as different line colors) from the native-like binding position at the D1 domain (after best superposition of the D1 domain on the reference structure). (C) RMSD of the protein backbone during the MD-simulation from the open (start) structure (pdb:4KR5) and from the closed state (pdb:4KQP). (D) Close-up view of a snapshot with bound Gln at the binding site superimposed on the crystal structure (shown in blue stick representation). The $\alpha$-carboxyl group of L-glutamine interacting with the guanidinium group of Arg333, forming a salt bridge, plays the major role in stabilizing this intermediate state.

small relative to the open start structure and much larger relative to the closed experimental structure (Figure 5.2 C). Also no other significant change in the RMSD of the protein upon ligand binding was observed. Hence, the GLN-binding to the specific binding-site on the larger D1 domain of SBD2 does not induce a closing of the two domains to directly form the closed bound structure. This indicates a significant energy barrier for closing that is not crossed on our simulation time scale. Since the simulation indicates a new potential intermediate SBD2 complexed state adopted prior to the domain closing event, it is of interest to characterize the binding process and how it is stabilized in detail. For the time intervals with a site-specifically bound GLN we performed an MMPBSA trajectory analysis to estimate the mean interaction contributions of individual residues in the SBD2 to stabilize GLN binding (Table 5.1).

| Residue | $\Delta$G[kcal/mol] |
|---|---|
| Arg333 | -6.3 $\pm$ 0.8 |
| Gly326 | -2.3 $\pm$ 0.4 |
| Ser325 | -1.7 $\pm$ 0.5 |
| Asp267 | -0.3 $\pm$ 0.5 |
| Ser328 | -2.6 $\pm$ 1.0 |
| Met327 | -2.2 $\pm$ 0.4 |
| Phe270 | -2.1 $\pm$ 0.6 |
| Phe308 | -2.5 $\pm$ 0.5 |

**Tab. 5.1.** Per residue interaction energy decomposition of the open SBD-Gln complexed state.

The calculated total MMPBSA binding energy was -9.5 kcal·mol$^{-1}$ (including a quasi-harmonic entropy term). The analysis of individual contributions revealed Arg333 to play the dominant role in stabilizing GLN binding by a salt bridge of the terminal COO- through the guanidinium group of Arg333. The backbone carbonyl

of Gly326 as well as the hydroxyl group of Ser328 also contribute significantly to ligand binding by forming frequent hydrogen bonds to the GLN alpha-amino group. The bound GLN side chain amino group forms stabilizing hydrogen bonds with Asp267 or Ser325 as well as non-polar interactions with the two opposing benzyl groups of Phe270 and Phe308. Assuming sufficient sampling of the underlying thermodynamic ensemble via this extended free MD simulation we determined the overall free energy difference of GLN binding in a straightforward fashion, by directly counting the fraction of simulation time spent in the respective states. The total simulation time amounted to 3384 ns in which GLN was bound to SBD2 for 2270 ns, leading to probabilities of Pbound= 0.67 and Punbound = 0.33. With a ligand concentration of Cl= 25mM a standard binding free energy of GLN binding of $\Delta$ G = -2.6kcal·mol$^{-1}$ was obtained.

In order to further analyze the dynamics and intermediates of the ligand binding process we investigated each association and dissociation event at higher time resolution. We recorded three distances in distinct regions of the binding site that characterize the bound state (see Figure 5.3). No distinct order of contact formation for the different events was observed. In intermediate states the GLN contacts to Arg333 and Gly326 are sharply peaked around the placements found also in the crystal structure of the bound complex and the formation as well as disruption occurs almost simultaneously. In contrast, the distance of the substrate to Asp267 shows a broader distribution with occasional higher deviations from the arrangement in the crystal structure. However, in some binding events it is the first contact formed and also the first that is disrupted upon dissociation.

### 5.3.2 Molecular Dynamics Simulations Starting from the Closed SBD2 Structure

Simulations starting from the closed conformation (PDB: 4KQP) with the bound GLN in the binding cleft between D1 and D2 domains indicated stable binding and also no tendency for global opening on the time scale of 1 $\mu$s (SI Figure C.2). However, the removal of the ligand (in silico replacement and solvation) from the closed crystal structure appears to destabilize the closed domain arrangement of SBD2 (see Figure 5.4). During simulations the closed state opened up within less than 100 ns and adopts a global open conformation in close agreement with the experimentally observed apo structure (Figure 5.4). The structures of the subdomains D1 and D2 do not change significantly upon opening with the RMSD

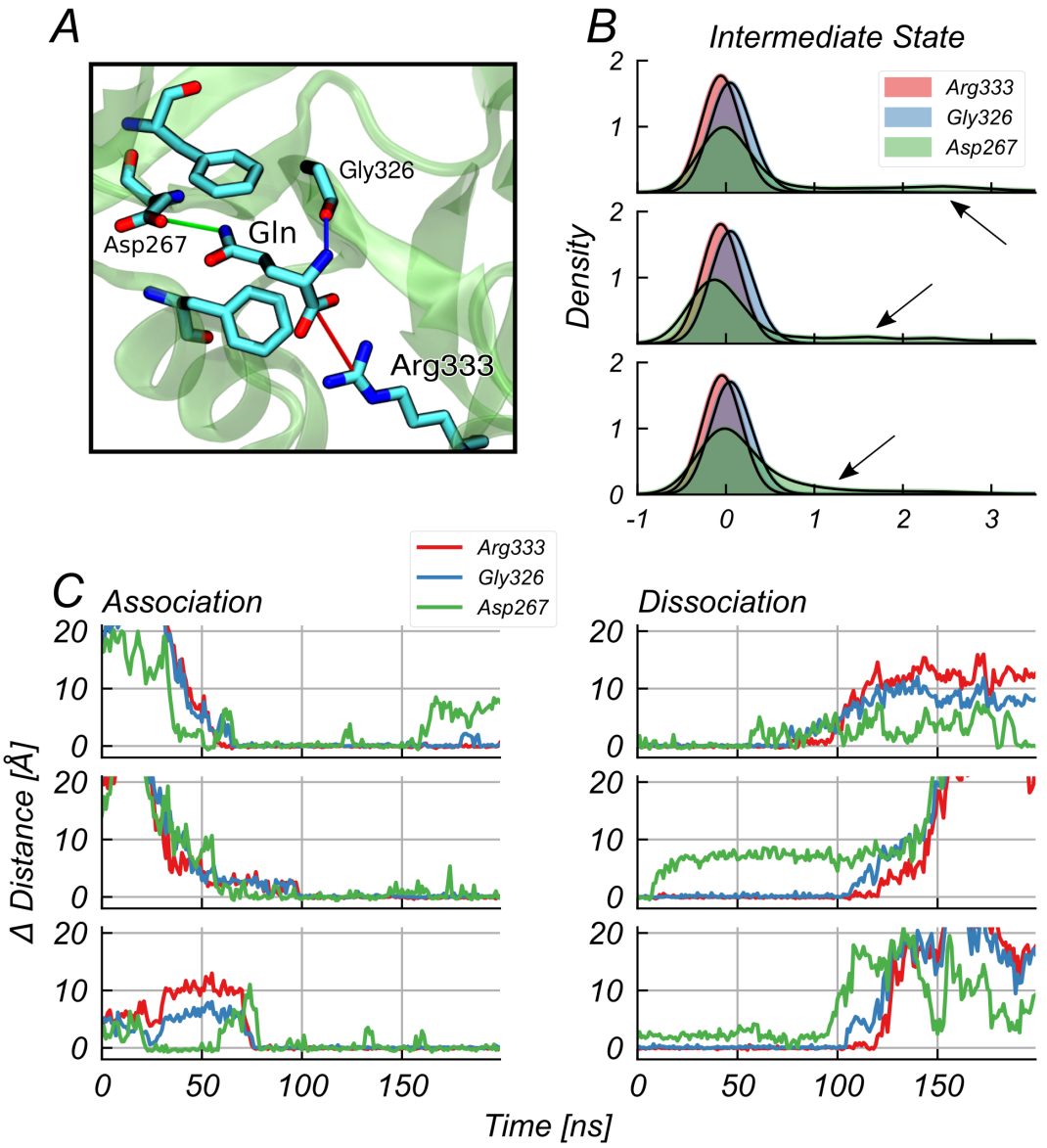



**Fig. 5.3.** Characterization of GLN ligand association and dissociation (with SBD2-D1) at high time resolution. (A) The overall position and orientation of GLN in the binding site on the D1 domain can be characterized by three contact distances in the ligand-bound crystal structure: The distance between Arg333:CZ and GLN:C atom (red line), Gly326:O and GLN:N atom (blue line) and Asp267:CG and GLN:NE atom (green line). The sampled distance-differences with respect to these reference distances in the crystal structure were used to capture the binding dynamics of GLN. (B-C) Sampled distance distributions and time resolved distance sampling for three association and three dissociation events of GLN during the MD simulation. In the intermediate state (liganded, but open protein form) GLN's amino group occasionally loses contact to the binding pocket (Asp267, Ser325) leaving the ligand solely stabilized by its charged functional groups and the hydrophobic contacts (high distance differences for Asp267: black arrows). Also, when examining the association/dissociation for the binding events, especially for the dissociation the contact to the Arg333 is the one firmly stabilizing GLN.

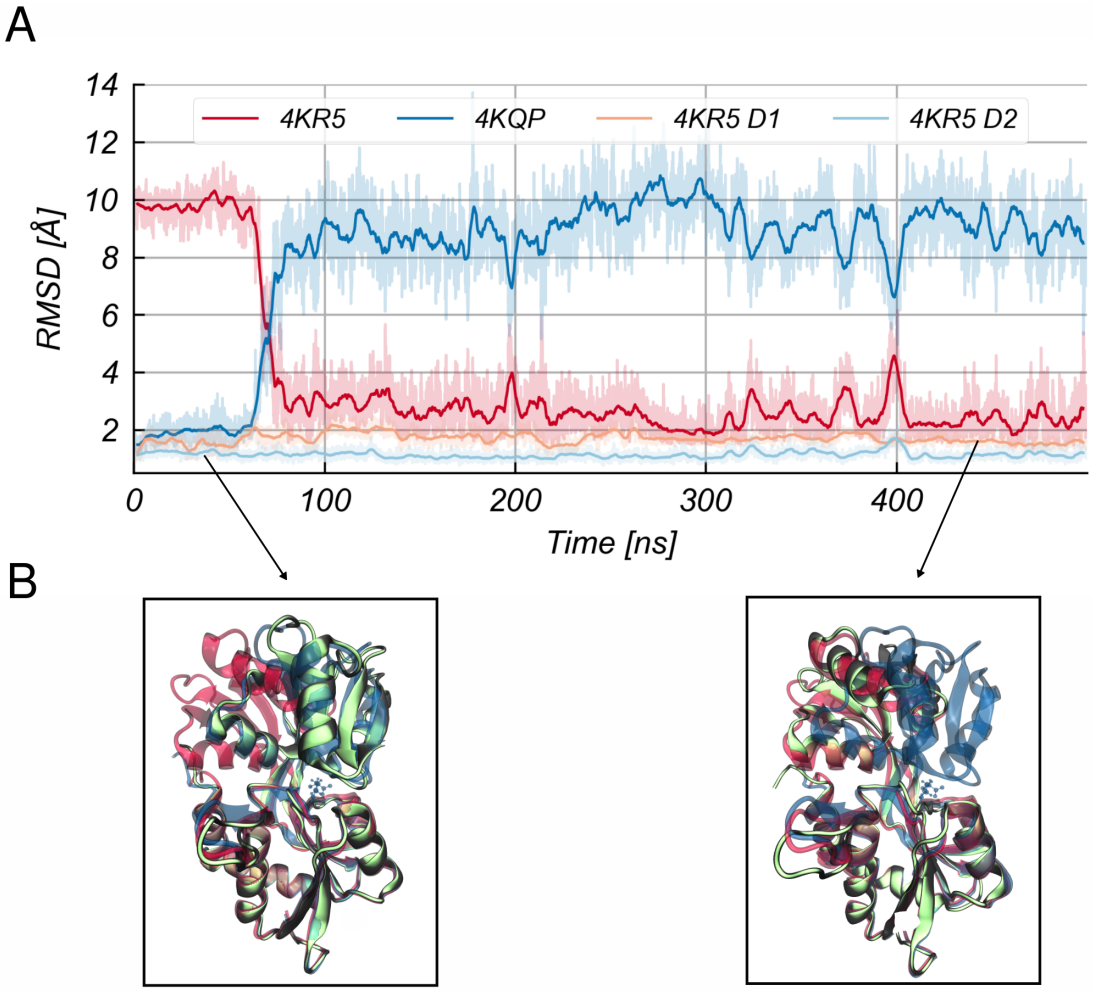



**Fig. 5.4.** Large-scale opening transition from the closed to the open conformation in the absence of ligand. (A) $C_{alpha}$- RMSD with respect to the crystal structures of the closed (PDB:4KQP, shown in blue) and open (PDB:4KR5, shown in red) conformations. The RMSD of the individual two domains D1 and D2 from the corresponding segments in the experimental start structure are also shown (yellow and light-blue lines, respectively). (B) Superposition of snapshots (with respect to D1 domain) during the simulation (green cartoon) onto the closed native structure (blue) and open native structure (red).

not exceeding 2-3 Å (similar to the observation when starting directly from the open SBD2, see above).

The most significant local structural rearrangements take place in the $\beta$-sheet hinge region. Backbone dihedrals in this region showed small fluctuations but mostly uncorrelated with respect to the closed or open SBD2 structure. However, we identified two correlated dihedral angles, the $\phi$-angle of Ser346 and the $\psi$-angle of Gly441, that allowed a separation between open and closed SBD2 (Figure 5.5, see also Figure C.3). It appears that a correlated dihedral change of just these two backbone dihedral angle changes mediates the global domain rearrangement (of course, more subtle changes of other dihedral changes may also contribute).

As a next step we applied principal component analysis (PCA) to identify the correlation in the backbone dynamics of this opening transition and projected the trajectory onto the first two eigenvectors (Figure 5.6, SI Figure C.4). That accounts for almost 95% of the observed global conformational change (Figure 5.6 C). The covariance matrix was calculated with respect to the non-hydrogen backbone atoms CA, C, O and N. The main feature of the transition is the ordered shift of the two domains D1, D2 towards the open state and the PC1 direction also correlates well with center of mass- distance between D1 and D2 (Figure 5.6 B). Note, that the C-terminal residues move towards D2's helix$_{418-427}$, especially in the realm of the key residue Leu480, which we will discuss in more detail subsequently (Figure 5.6 B).

The relative changes in root-mean-square-fluctuation (RMSF) of each residue gives a measure of residues that showed different mobility in the two domain arrangements (Figure 5.7). Typically, ligand binding to a receptor leads to a reduction of conformational fluctuations and we also observe that the mobility of most residues is slightly smaller in the closed vs. open states. Surprisingly, for several mostly hydrophobic residues the mobility is dramatically reduced in the open state compared to the closed form of SBD2 (Figure 5.7 A). One can interpret this as an indicator of stabilizing interactions, especially visible for the C-terminal (and D2 $\alpha$-helical$_{418-427}$) region when the protein is in an open state. All the residues with reduced mobility in the absence of GLN (Pro272, Ile285, Tyr344, Pro419, Leu480) contribute to the formation of a hydrophobic 'lock', that stabilizes the open state of SBD2 (Figure 5.7 B-D). Note, the spontaneously formed arrangement (not seen in the closed structure) observed during the unrestrained simulations starting from the closed form is in close agreement with the local arrangement found in the experimental structure of the open form (SI, Figure C.5).

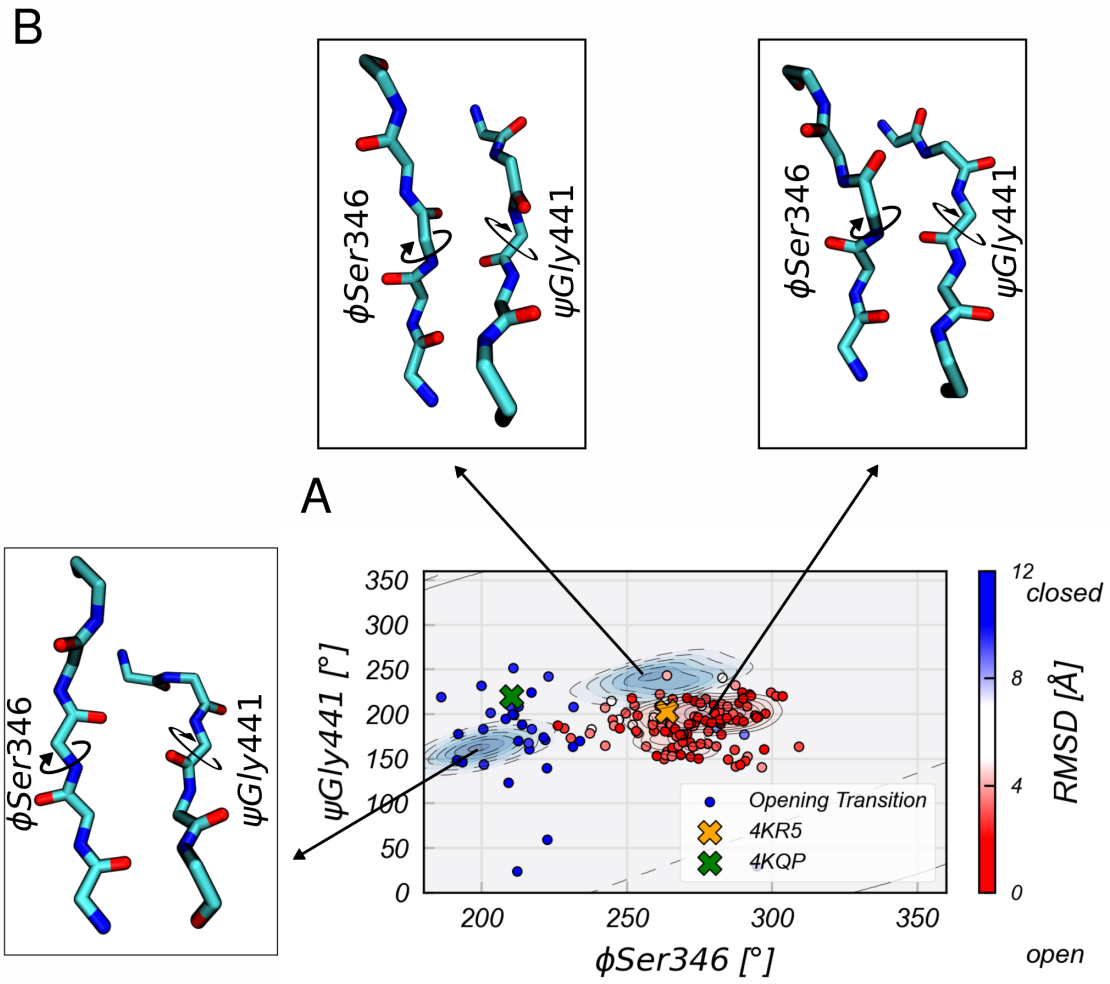



**Fig. 5.5.** Conformational changes in the $\beta$-sheet hinge region upon opening transition of SBD2. (A) The $\phi$-Ser346 and $\psi$-Gly441 dihedral angle distributions are depicted as blue and red contours extract from the simulations of the open (red) and closed (blue) SBD2 structures. In addition, the states sampled during the transition from a closed to an open conformation (observed in the time interval of 60-90 ns as illustrated in Figure 5.4) are indicated by blue to red dots (color scaled by deviation from closed vs. open SBD2 structure). The reference dihedral angles of the respective crystal structures are shown as crosses (orange, green). (B) Conformational snapshots of the hinge region that correspond to the dominant states are illustrated as stick models.

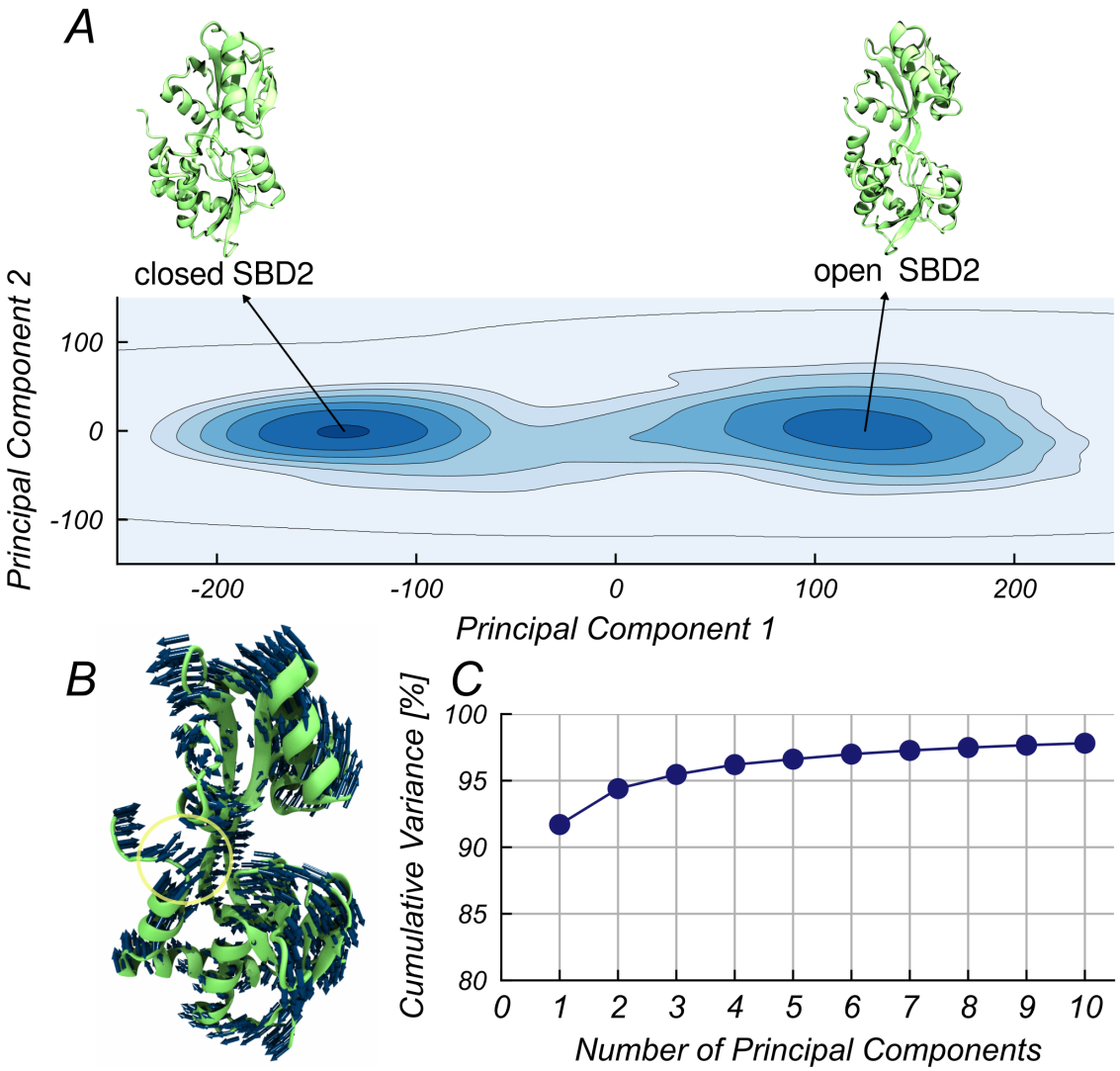



**Fig. 5.6.** Principal Component Analysis of the opening motion, upon in silico removal of Gln. (A) Projection of the sampled transition onto the first two eigenvectors. (B) Illustration of the collective backbone atom directions of the first dominant PC (contributing 92% of the variance). Note, that in the first component the transition of the two subdomains away from each other is correlated with a second collective movement of the D1 C-terminal$_{417-484}$ region towards D2's $\alpha$-helix$_{418-427}$ . (C) Cumulative contribution of the first 10 PCs to the total Cartesian variance sampled during the simulations.

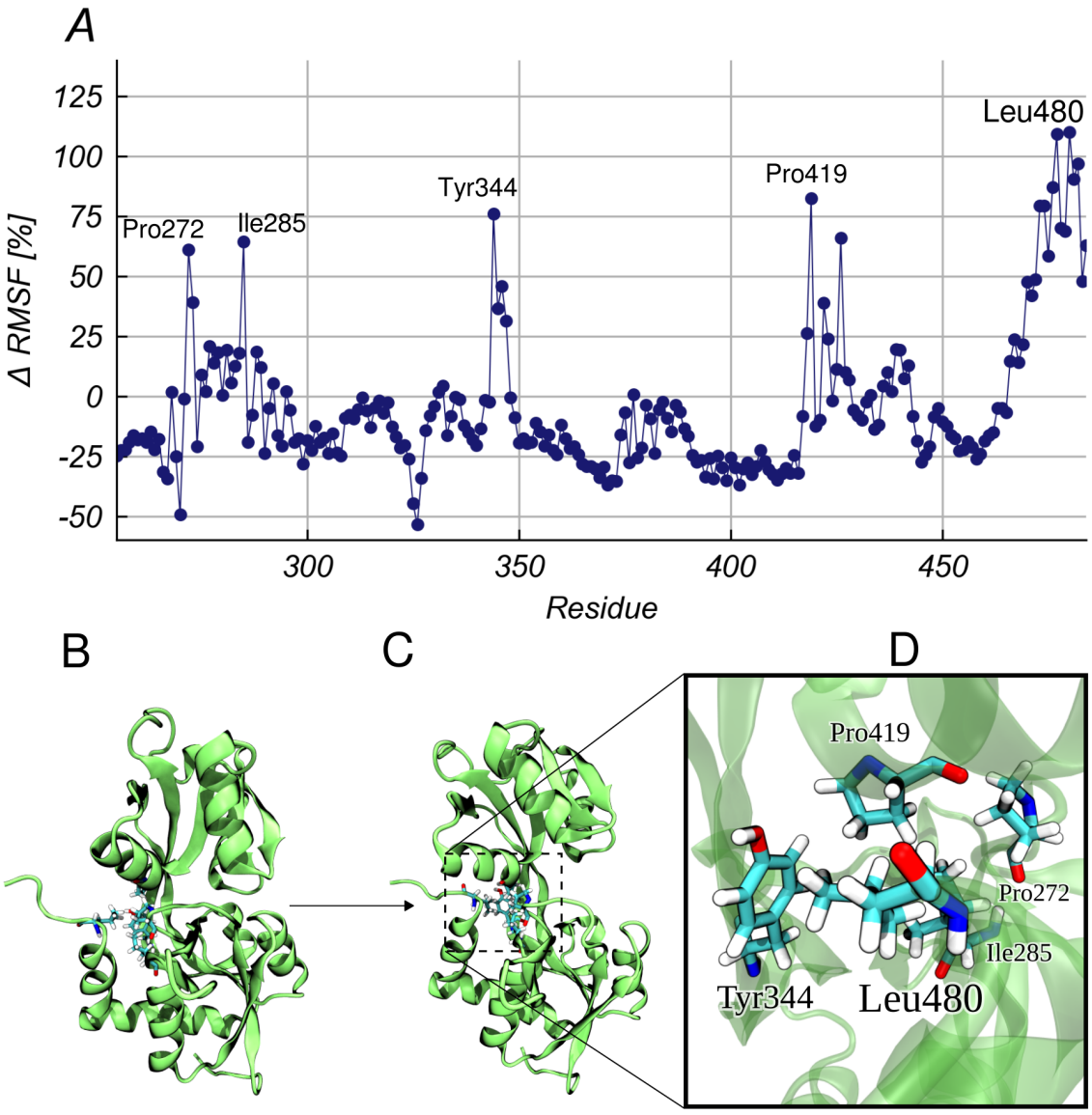



**Fig. 5.7.** (A) Relative $\delta$ RMSF per residue in SBD2 upon opening of the protein. A positive/negative $\delta$ RMSF indicates a higher mobility in the closed/open form, respectively. The few residues with positive $\delta$ RMSF indicate residues that participate in the D1-helix$_{471-484}$ –D2-helix$_{418-427}$ interaction. (B, C) Illustration of the formation of D1-helix$_{471-484}$ –D2-helix$_{418-427}$ contacts upon SBD2 opening. (D) A close-up view of the interacting residues.

### 5.3.3 Effect of the Mutation Leu480Ala on the SBD2 Dynamics

The above unrestrained simulations starting from the closed SBD2 structure without ligand indicate that the C-tail$_{471-484}$ residues of D1 form an interaction with D2's $\alpha$-helical$_{418-427}$ residues, to serve as a 'lock', stabilizing the open conformation. Despite the fact that other D1 residues (Pro272, Ile285) as well as tyrosine (Tyr344) in the $\beta$-sheet hinge region also participate in this hydrophobic 'locking' mechanism it is this close contact between the two helices which seems to be crucial in the open SBD2 state. Interestingly, this type of D1/D2 contact has been found in several SBDs (SI Table C.1) and is characteristic for the SBD2 class of ligand binding proteins. Especially, the Leu480Ala residue is centrally located at the hydrophobic pocket formed by residues of the D1 domain (Figure 5.7 D) and shows also a much lower local mobility in the open vs. closed SBD2 structure (Figure 5.7 A). The substitution of this residue by alanine is expected to lower the interaction between the two lobes of SBD2 and should reduce the energy barrier/penalty for global domain rearrangements. Hence, comparative MD simulations were started from the open structure of the SBD2 with the Leu480Ala in silico substitution either in the presence of a GLN ligand bound to the D1 domain (same geometry as observed in our simulations) or without a ligand (Figure 5.8). Already after 220 ns of unrestrained MD simulation with GLN present the SBD2 undergoes a global transition from the open state towards the closed D1/D2 geometry with an RMSD of $<$ 2.5 Å with respect to the experimental closed structure (compared to no closing event within $>$ 3000 ns for the wild-type complex, Figure 5.2). During this transition the conformation of the two individual D1, D2 subdomains remains close to the conformation in the crystal structure with an average RMSDs values of $\sim$3 Å for D1 and 2.9 Å for D2. Again, the closing motion is controlled by backbone dihedral angle changes in the $\beta$-sheet hinge region connecting the two lobes, with correlated changes in $\phi$-Ser346 and $\psi$-Gly441 separating the two states (SI Figure C.3). After closing, the simulated structure remains close to the experimental crystal structure (Figure 5.8). Interestingly, even in the absence of a ligand a spontaneous transition to the closed from was observed (at around 1650 ns). This transition also resulted in the disruption of the contact between C-tail$_{417-484}$ and the $\alpha$-helix$_{418-427}$ of D2. However, the closed form showed larger conformational fluctuations than in the presence of a bound ligand (Figure 5.8 A) and the lifetime of the state was only $\sim$350 ns before it returned to the open SBD2 arrangement. All discussed transitions from open to closed protein conformation and vice versa seemed to be a clean tilting motion, be it with or without ligand. To verify this, we

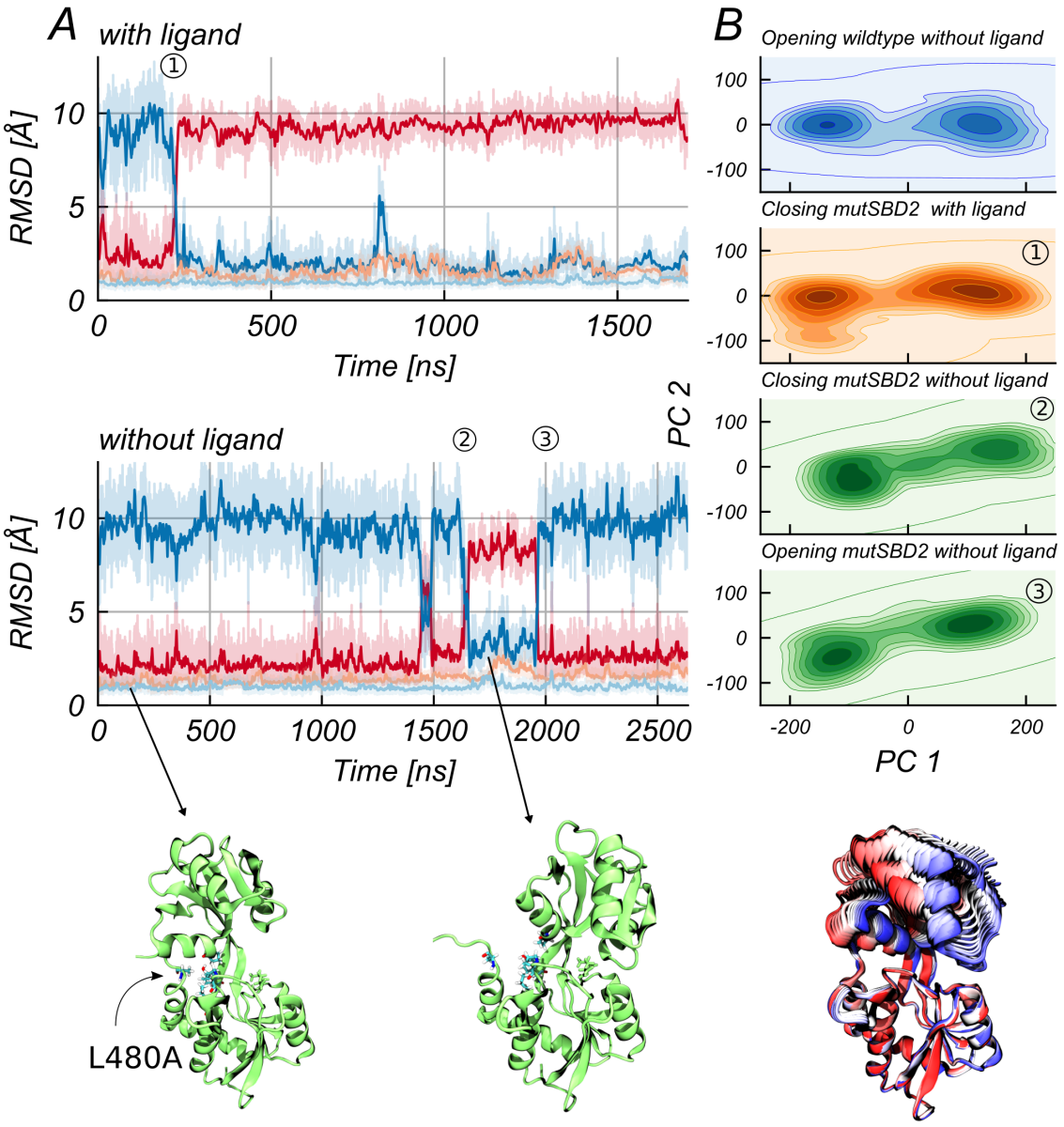



**Fig. 5.8.** SBD2-L480A reduces the open-close transition barrier. (A) RMSD of sampled states from the native open (red) and closed (blue) states of SBD2 during simulations. The RMSD of individual domains D1 and D2 relative to the native structures is illustrated by yellow and light blue lines. Note, that the presence of the ligand (lower panel in (A) drastically increases the lifetime of the closed state. Typical sampled snapshots of the open and closed state are shown as cartoons below the plot. (B) Projection of the sampled states from the SBD2-L480A transitions onto the first two PC eigenvectors of the opening trajectory for the wild type protein. The height of the density is scaled logarithmically. The two sampled clusters represent the two stable protein states. All transitions followed the same pattern, namely a clear tilting motion. Multiple frames of a typical transition trajectory (lower panel in B) illustrate this tilting motion (shown in cartoons, colored from beginning (red) to end of transition (blue)).

projected all the transitions for SBD2/Leu480Ala onto the first two eigenvectors for the wild type closed-to-open transition (scaling the density logarithmically, Figure 8B). The first eigenvector represents the global motion, clearly separating the two states of the protein. All transitions displayed a clear tilting motion, with disruption of the D1/D2 contact (lower panel Figure 5.8 B).

### 5.3.4 Free Energy Advanced Sampling Simulations of SBD2 Domain Motions

The unrestrained MD simulations give important qualitative insights into the substrate dependent local and global dynamics of the SBD2 system. However, important transitions such as the open-to-close transition of the SBD2 protein upon ligand binding was only observed for a Leu480Ala mutation but not for the wild type indicating a substantial energy barrier not crossed within the simulated time scales. In order to investigate the free energy profile for open-close domain transitions we performed extensive HREUS simulations on the system employing a center-of-mass distance reaction coordinate between the two subdomains D1 and D2 that allows to distinguish the open and closed state and correlated with the PC1 of the observed transition. The HREUS free energy simulations were performed in the presence and absence of GLN in the binding cleft on the D2 domain. The frequent exchanges between Umbrella sampling windows in the HREUS approach result in good convergence of the calculated free energy profiles (Figure 5.9) but still requiring extensive sampling of $\sim$ 500 ns per US interval (aggregate US simulation $>$ 30 $\mu$s). In the apo state the open domain arrangement is strongly favored by a free energy difference of $\sim$3.9 kcal·mol$^{-1}$. Only a small energy barrier of $\sim$0.1-0.2 kcal·mol$^{-1}$ between closed state and transition towards opening was observed. In parallel to the induced closing or opening the C-tail$_{471-484}$-D2 helix$_{418-427}$ contact is formed/disrupted (Figure 5.9 B). In the presence of a GLN ligand in the binding pocket the free energy of the closed state is found to be slightly lower (by about $\sim$ -0.3 kcal·mol$^{-1}$) compared to the open state. The barrier for open-to-close transition is similar to the barrier in the opposite direction and amounts to $\sim$2.9 kcal·mol$^{-1}$ (Figure 5.9). Interestingly, the formation/disruption of the additional C-tail$_{471-484}$-D2 helix$_{418-427}$ contact coincides with the transition barrier indicating that this is the key-element for the observed conformational switching and distinction of two stable states of the system. Combined with the GLN binding in the unrestrained simulations it is possible to suggest a thermodynamic model for the

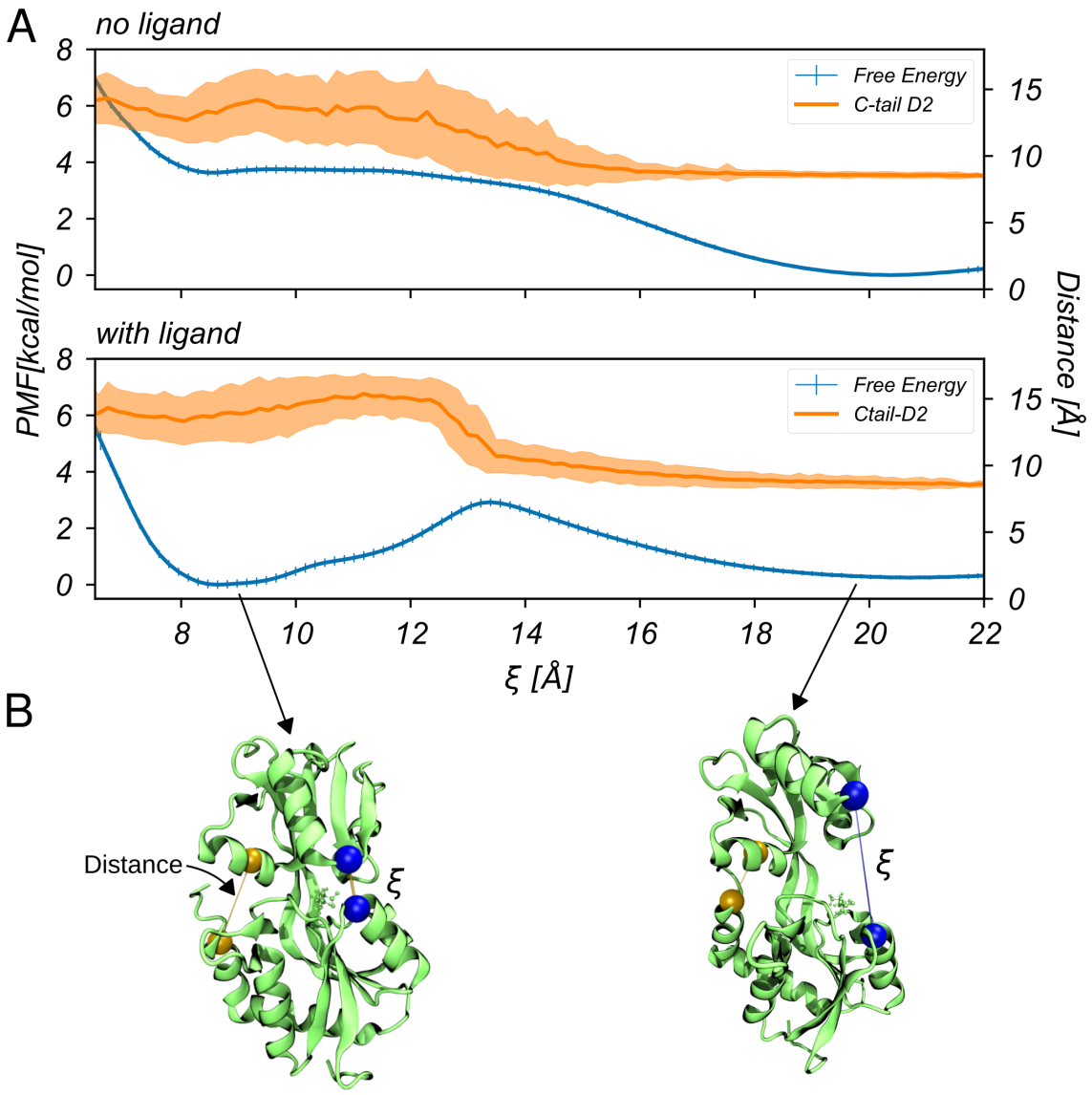



**Fig. 5.9.** (A) Computed potentials of mean force (PMF) for the wild-type SBD2 vs a center-of-mass distance coordinate between centers indicated as blue spheres in the protein structures (B) (see also Methods and SI, Figure C.6). The regime around $\xi$=8.5 Å represents the closed state and around 19-20 Å the open SBD2 state. The presence of L-glutamine in the binding site (lower PMF panel) stabilizes the closed state. In addition, the distance of the C-terminal D1-helix$_{471-484}$ – D2-helix$_{418-427}$ is illustrated. A small distance and small fluctuations indicates a stable contact with little fluctuations (in the open state) and larger distances and stronger fluctuations are observed for the regime representing the closed state. The mean values of this distance together with the standard deviation are shown.

binding and coupled conformational switching process (see Figure 5.10). The total calculated binding free energy of $\sim$3 kcal·mol$^{-1}$ suggests only weak binding in the mM range compared to a Kd in the $\mu$M regime measured experimentally[208, 217]. In the absence of the GLN ligand the open form is favored and transition to the closed form involves a significant transition penalty and barrier. The GLN ligand has already significant affinity to the open form (binding only to the D1 domain). The calculated binding free energy for this step can only be considered as estimate because it was derived from unrestrained MD simulations with only few bound/unbound transitions. Surprisingly, the transition to the closed form also involves a free energy barrier (such that a spontaneous transition in the unrestrained simulations was not observed) and results only in a small further stabilization of the GLN-bound form.

## 5.4 Conclusions

The SBD2 protein is an important subunit of the GlnPQ transporter system and binds specifically GLN molecules in order to deliver it to the translocation system of the GlnPQ transporter. It has been investigated extensively using biochemical experiments, NMR spectroscopy [219], X-ray crystallography [208] and time-resolved smFRET experiments [204, 217, 218]. The structural as well as smFRET studies demonstrate a global conformational transition coupled to ligand binding from an open apo to a closed domain arrangement upon ligand binding. The smFRET experiments indicate occasional domain closing events in the absence of GLN molecules that increase significantly in frequency upon addition of GLN ligands [218].

In the present study our MD simulation results indicate that GLN ligands can bind to SBD2 site-specifically with low affinity at exactly the ligand binding position found experimentally on just the D1 domain without an immediate domain closing motion. This indicates that ligand binding and domain motion do not occur simultaneously as expected in an induced fit binding process. However, HREUS simulations clearly indicate that a bound ligand strongly stabilizes the closed arrangement but the transition from and open to a closed domain state involves a significant energy barrier that is not crossed on the time scale of our unrestrained MD simulations. In smFRET experiments the binding of the GLN ligand itself cannot be detected and only an increased tendency of closing in the presence of GLN is observed [218]. However, smFRET experiments with added of L-Arginine prevents GLN induced closing of SBD2 [217]. L-Arginine binds to SBD2 and at sufficient concentration

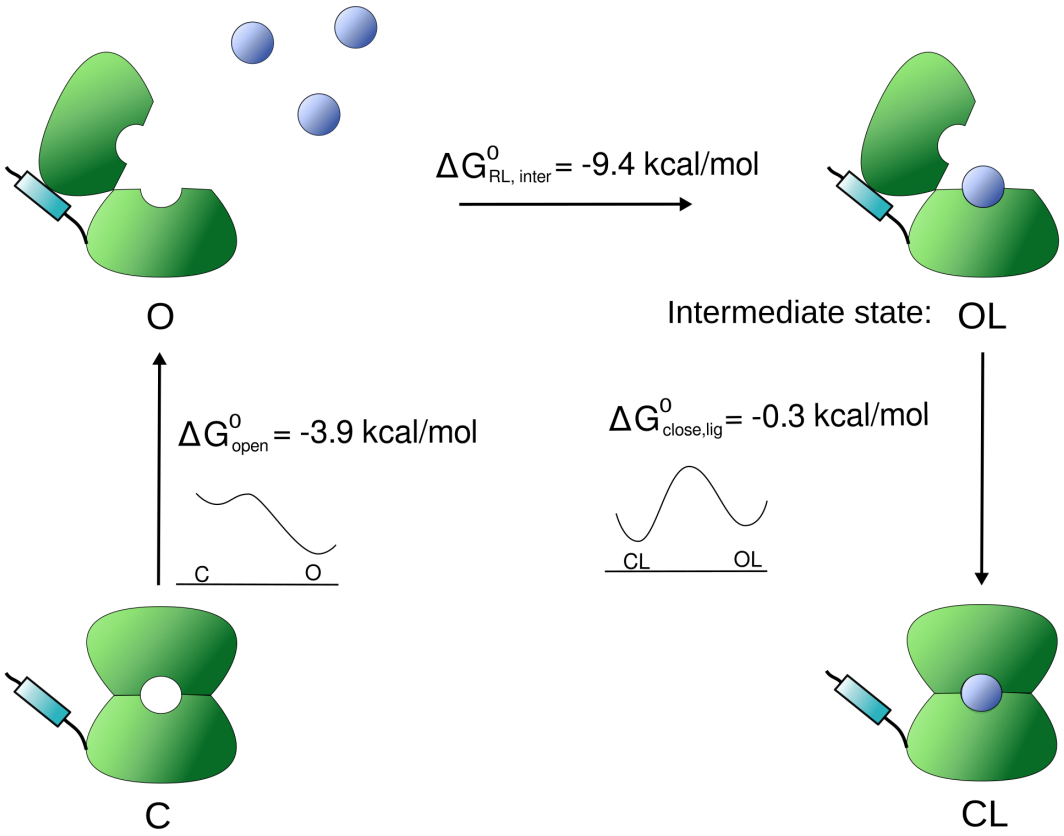



**Fig. 5.10.** Ligand-binding mechanisms and SBD2's conformational changes: The open protein form is locked via hydrophobic interactions between the subdomains in the C-terminal region (highlighted by blue rectangles). SBD2 recognizes its substrate via a two step mechanism: Glutamine binds to the active site of the larger subdomain D1 of SBD2 in open conformation, forming an intermediate state. The ligand then induces a transition to the closed state by stabilizing the closed conformation of SBD2.

competes with GLN (hence binds at the same pocket of region) but does not induce closing. This supports the present result that binding and closing do not occur simultaneously (not an induced fit binding). L-Arginine binds at the same pocket as GLN on the D1-domain but due to its positive charge it interacts with the D2 domain less favorable than GLN (it repels it because of the positive charge, D2 is also positively charged). It is important to note that the present major finding of a pre-equilibrium of GLN binding to just the D1 domain and only a small stabilizing contribution upon closing makes sense from a biological view point: GLN needs to be bound with modest affinity but for the transport it is important that the free energy cost of domain opening to release GLN is small to minimize the dissociation of the SBD2 complex before releasing the substrate to the ABC transporter.

Another major result of the present study is the observation that the D1-tail$_{471-484}$ –D2-helix$_{418-427}$ interaction is a decisive element to stabilize the open state and it controls the barrier for the open-closed transition. This interaction element is typical for the SBD2 class of periplasmic ligand binding proteins and it is likely that it plays also in these cases a similar key role as in case of SBD2. The weakening or disruption of the contact (by the Leu480Ala substitution) resulted in rapid (barrier less) closing and it also destabilizes the open state such that occasionally closed states are sampled in the absence of ligand on the present MD timescale. FRET-experiments on the SBDLeu480Ala mutation indeed confirmed these findings with an increased population of the closed form already in the absence of substrate [229]. The barrier for the transition from the closed SBD2 state to the open state is found to be significant in case of a bound GLN but considerably smaller for SBD2 without a ligand. Hence, the prediction from this result is that the life time of the closed form should be longer in the presence of a ligand than without a ligand. However, smFRET studies show that addition of GLN increases the frequency for open-to-close transitions [218]. This agrees qualitatively with the simulation results because the barrier height in the absence of ligand is larger than in its presence in the binding cleft but the life time of the closed state was found to be similar both in the presence and absence of ligand [218]. Such scenario predicts a similar barrier height for the global opening process in the presence and absence of ligand which was not observed in the simulations. It is possible that differences in the diffusivity along the open-closing pathway in the presence or absence of GLN may play a role. However, it should be kept in mind that due to limited force field accuracy a quantitative agreement with all experimental results is not expected. Future simulation studies could give important insight into similarities and differences of the molecular substrate binding mechanism of related periplasmic binding proteins.

# 6

# The Inhibition of Substrate Binding Domain 2 by Arginine [1]

> *Hofstadter's Law: It always takes longer than you expect, even when you take into account Hofstadter's Law.*
>
> **— Douglas Hofstadter**

Substrate-binding proteins (SBPs) are of central importance for the uptake of vital nutrients into the cell. However, the precise relationship between the mechanism of ligand binding and the consequent conformational changes within domains remains unclear for many SBPs. Our study aims to address this intricate issue, specifically concerning the Substrate Binding Domain 2 (SBD2) of the bacterial glutamine (Gln) transporter from *Lactococcus lactis*. SBD2 plays a pivotal role in facilitating the uptake of glutamine. In this particular system, single-molecule Förster resonance energy transfer (smFRET) measurements have revealed that L-arginine possesses the ability to competitively inhibit Gln from triggering the closed conformation of SBD2. Moreover, L-arginine impedes the translocation of Gln into the cell mediated by SBD2. Using extensive molecular dynamics simulations, we have uncovered the precise binding mode of arginine to SBD2 in atomistic detail. Our findings reveal that L-arginine establishes similar contacts with SBD2's larger subdomain D1, akin to its interaction with the cognate substrate, Gln. Furthermore, our analysis has identified specific interactions that impede the transition to the closed state when arginine binds to SBD2. Notably, Coulombic repulsion between the charged arginine and Lys373 emerges as a critical interaction underlying this inhibition. The proposed SBD2(K373A) mutation shows promise as a potential tool to validate whether a conformational mismatch between open SBD2 and the translocator is responsible for hindering arginine uptake into the cell.

[1] Apart from minor modifications, this chapter was originally published in [230]

## 6.1 Introduction

The survival of cells depends on the selective uptake of nutrients from their surrounding environment. Periplasmic substrate (or solute)-binding proteins (SBPs) play a central role in such processes as they mediate the translocation across the membrane [209, 211]. Their capability to bind a diverse array of molecules, encompassing ions, vitamins, sugars, peptides, amino acids and inorganic anions, underscores their pivotal role in cellular function [210]. Structurally, SBPs are characterized by two rigid subdomains (D1 and D2) bridged by a flexible $\beta$-sheet hinge region [208]. Classification of these proteins into different structural types is predicated upon their core topology and the composition of this $\beta$-sheet region. For instance, type 1 SBPs feature three strands interconnecting the domains, while type 2 SBPs exhibit only two beta strands. Further categorization into distinct clusters is based on their structural folds and substrate affinities [231]. Numerous crystal structures suggest the existence of many SBPs in either a ligand-free (apo) open conformation or a ligand-bound (holo) closed state. The conformational states and their associated dynamics are pivotal determinants in the overall transport process [232, 233]. Various biophysical techniques have been employed to unravel mechanistic insights into ligand binding and domain movement of SBPs [218, 234–236].

However, establishing a precise correlation between the mechanism of ligand binding and the conformational movement of domains in SBPs is a challenging endeavour. For instance, certain SBPs undergo a conformational shift to the closed form upon binding of different ligands, or even in the absence of any ligand [216, 237]. Conversely, in SBPs such as MalE, binding of non-cognate substrates appears to impede the transition to the closed state [238, 239]. Single molecule FRET (smFRET) experiments have indicated that some of these proteins exhibit extensive plasticity, suggesting the presence of multiple semi-closed conformations that are not easily categorised as open or closed states [217]. The extent to which ligand binding precedes domain closure (induced fit model) or whether conformational changes can occur independently of prior ligand binding (conformational selection model) has been investigated for several SBDs using sFRET, suggesting that an induced-fit mechanism is involved [218, 240–242]. However, conflicting findings regarding the coupling between ligand binding and overall adaptation in the L-glutamine (GLN) binding protein GlnBP from E. coli have been related to the challenge of isolating one mechanism from another through experimental approaches [243]. Moreover, within this diverse array of binding proteins, for-

mulating a universal model is not only challenging but often deemed impossible due to subtle sequence disparities resulting in distinct and unique structural and dynamic behaviors[231]. Consequently, it appears inevitable that understanding must be garnered for each substrate-binding domain (SBD) on an individual basis and utmost caution is necessary when attempting to draw generalizations across these diverse proteins.
Interestingly, substrate binding and transport can be disrupted or even prevented by competing binders. This can involve non-cognate ligands, which can act as an inhibitor for certain bacterial transporters. For example, the SBD2 of GlnBP binds GLN, which is subsequently efficiently transported. However, the process can be inhibited by arginine (ARG). Experimental evidence has demonstrated that ARG not only competitively inhibits GLN binding and associated conformational changes induced by GLN, but also impedes the uptake of GLN into the cell [217]. This observation suggests that the non-cognate ARG also binds to SBD2 but fails to induce a conformational change. Understanding the inhibition by ARG in molecular detail could provide valuable insights into the mechanism of substrate binding and associated conformational changes necessary for substrate transport.

In this chapter, we employed molecular dynamics (MD) simulations focusing on the substrate-binding domain 2 (SBD2) of the GlnPQ importer derived from the gram-positive bacterium *Lactococcus lactis*. The objective of this study was to gain a deeper understanding of the intricate relationship between ligand binding and the overall adaptation mechanism within this type 2 SBD. The high temporal resolution of MD simulations enabled the separation of ligand binding from domain motion in this system, thus providing a valuable complement to other experimental techniques.

The present study employs extensive continuous (c)MD simulations and advanced sampling protocols to elucidate the binding mechanism of ARG to SBD2 in atomic detail and to compare it to that of its cognate substrate, GLN. In cMD simulations initiated from the open apo SBD2, reversible transient binding of ARG is observed to occur at the same site on the D1 domain that also binds GLN. This binding is not accompanied by closure within the MD time scale. The binding strength is characterised using free energy simulations and compared to that of GLN binding. We find that free energy simulations to induce closure indicate a significant penalty when ARG occupies the native binding site, in contrast to GLN, which favours closure. This result explains both the competition effect of ARG and offers an explanation for why transport is inhibited by ARG, which prevents closing. The analysis identifies the underlying interactions preventing the adaptation of the

closed state in the presence of arginine within the binding pocket. Based on the simulations, we also studied a K373A mutation, which we find to allow closure even in the presence of bound ARG. This discovery has significant implications, as it may help to elucidate whether a conformational mismatch between the open SBD2-ARG complex and the transmembrane domains (TMDs) of the GlnBP is responsible for the inability to transport ARG into the cell. It should be noted that analogous coupling mechanisms have been proposed for other SBDs. For instance, in the maltose importer MaFGK2, which is one of the most thoroughly studied importers, a "concerted" model of transport has been proposed. This model proposes that the maltose-bound MalE (SBD of MaFGK2) and ATP act in concert to initiate the catalytic cycle by binding to and reducing the energy of the intermediate conformation [244]. This process facilitates partial rotations in the TMDs, aligning critical catalytic residues at the NBD dimer interface. This, in turn, facilitates ATP binding and subsequent transport into the cell [245]. Moreover, it has been demonstrated that only the substrate-bound closed form of MalE efficiently stimulates ATPase activity [246]. However, it should be noted that a conformational mismatch might not be the only mechanism hampering ARG uptake. For example, the binding to SBD2 might be too stable to allow efficient release of the substrate to the translocator. In other importers, substrate binding pockets have been identified within the TMDs, which might also be linked to transport regulation [247, 248]. The proposed SBD2(K373A) mutation could potentially aid in deconstructing the intricate translocation process into discrete components for the SBD2 protein in the GlnPQ complex.

## 6.2 Methods

For all MD simulations, the AMBER20 software suite and the ff14SB force field were used [141, 142]. Water was treated explicitly with the TIP3P water model [143].

### 6.2.1 Unrestrained MD Simulations

Unrestrained MD simulations were started from the crystal structure of the open SBD2 conformation(PDB:4KR5) from Lactococcus lactis [208]. The protein was solvated in an octahedral box with a minimum distance of 10 Å to the box boundaries.

The protonation states of the titratable amino acids were predicted via Poisson-Boltzmann calculations, using PROPKA3 [249]. Six ARG molecules were placed randomly into the simulation box, resulting in a ligand concentration of ∼25mM. Sodium and chloride ions were added to neutralize the system and reach an ion concentration of 100 mM. After an energy minimization of 2000 steps of steepest decent, the systems were gradually heated up to 300K while keeping positional restraints (force constant of 6kcal $\cdot$ mol$^{-1}\cdot$ Å$^{-2}$) on the protein-non-hydrogen atoms during 1 ns simulation time. The restraints were gradually removed for another 1 ns followed by unrestrained production simulations of more than 6$\mu s$. All bonds involving hydrogen were kept at a optimal length using SHAKE [82]. The hydrogen mass repartition scheme was used allowing time steps of 4fs [83]. Trajectory analysis was performed using cpptraj and pytraj [153]. Visualization of structures and trajectories was performed using VMD [154].

### 6.2.2 MMPBSA Trajectory Evaluation

Average interaction energy calculations on the unbiased trajectories have been conducted using the MMPBSA trajectory post processing method, utilizing the MMPBSA.py program of the AMBER package.[97] [96] The purpose of the calculations was to obtain an estimate of the contribution of individual residues of protein on the binding of the ARG ligand. Here, the 'single trajectory approach' was used, assuming that there are no significant conformational changes upon binding and yielding the mean interaction between substrate and receptor decomposed for each residue at the binding region. For this, 3000 frames in the ligand-bound state of SBD2 sampled during free unrestrained MD simulation (Fig. 6.2 A) in the range from 100 to 1800 ns have been evaluated. The ion concentration was set to 100mM (same as in the explicit solvent simulations). The harmonic conformational entropy contribution was included for calculating the total binding energy.

### 6.2.3 Absolute Binding Free Energy Calculations

To quantitatively compare the binding affinities of both GLN and ARG to the open form of SBD2 a well-established, advanced sampling protocol, developed by H.-J. Woo and B. Roux was applied [250]. This approach effectively addresses convergence issues typically encountered in standard Umbrella sampling while

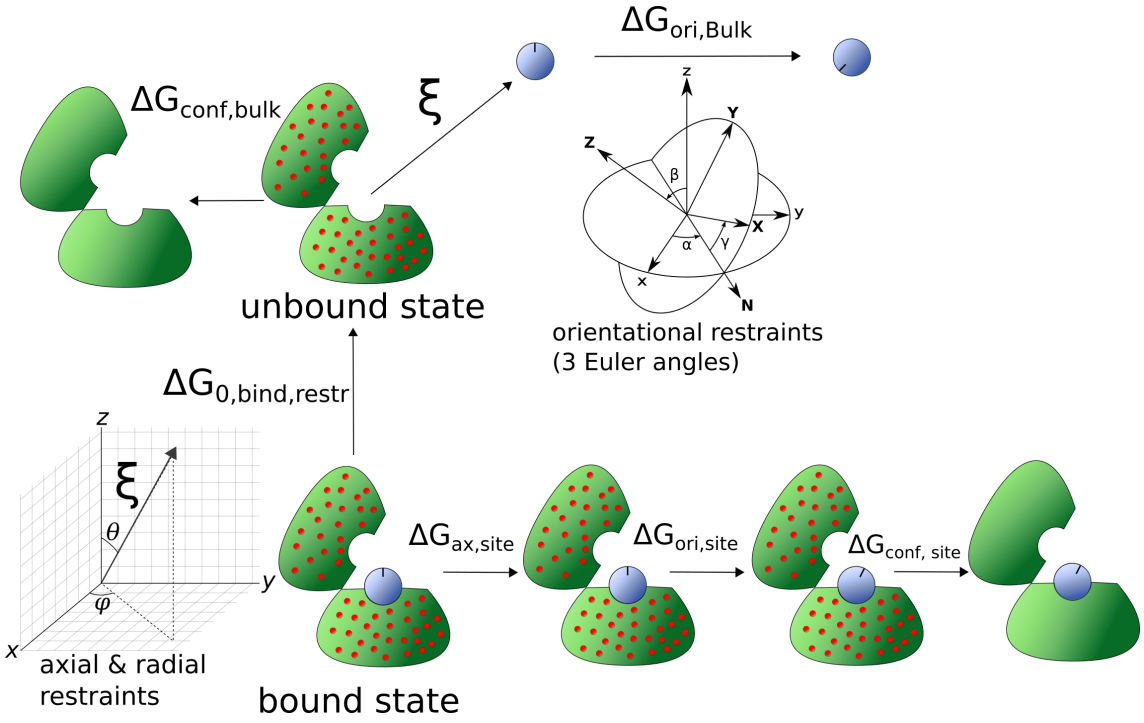



**Fig. 6.1.** Schematic illustration of the contributions to absolute binding free energy calculations. In addition to the PMF along the radial separation coordinate $\xi$ with applied axial, orientational and conformational restraints, the contributions associated with the release of these contributions have to be calculated in the bound and unbound state. The $C_\alpha$ atoms used to calculate the conformational RMSD restraint are indicated as red dots for illustrative purposes, demonstrating the inclusion of conformational restraints.

calculating binding free energies, substantially reducing the overall simulation time.

A one-dimensional potential of mean force (PMF) along the reaction coordinate $\xi$, defining a radial center-of-mass (COM) distance between ligand and receptor, is calculated. This calculation involves the imposition of restraints on several orthogonal degrees of freedom, along with maintaining conformational restraints on the binding partners.

The restraints were defined between three virtual sites in the open conformation of SBD2 and the according ligand of interest (Gln,Arg)(for exact definitions of the sites see SI, Figure D.1).

Due to the small sizes of both ligands conformational restraints were only applied on the SBD2 protein. The restraint binding free energy $\Delta G_{bind,0,restr}$ needs to be corrected to obtain a physically meaningful standard binding free energy $\Delta G_{bind,0}$. First, the introduction of axial and orientational restraints is corrected for via free energy perturbation at the binding site ($\Delta G_{a,site}$, $\Delta G_{o,site}$) and an analytical

calculation in the bulk ($\Delta G_{o,bulk}$). Note, that $\Delta G_{a,bulk}$ is accounted for implicitely in the context of $\Delta G_{bind,0,restr}$. Finally, the free energy contributions associated with the conformational restraints are obtained by Boltzmann-weighted integration of a PMF calculated along a RMSD coordinate. For the precise mathematical derivations, please refer to the original work of H.J. Woo and B. Roux [250].

The standard binding free energy is the sum of the individual contributions:

$$\Delta G_{bind,0} = \Delta G_{bind,0,restr} + (\Delta G_{o,bulk} + \Delta G_{c,bulk}) - (G_{a,site} + \Delta G_{o,site} + \Delta G_{c,site}) \quad (6.1)$$

The restraint PMF along the center of mass distance coordinate was calculated via Umbrella sampling by pulling the ligand molecule out of an binding position from the SBD2 binding site. The biasing potentials cover a distance range from 3 Å to 29.5 Å with a window spacing of 0.5 Å. Each window was simulated for 10 ns with a restraint force constant k of 15kcal $\cdot$ mol$^{-1}\cdot$ Å$^{-2}$ while keeping a force constant of 100kcal $\cdot$ mol$^{-1}\cdot$ rad$^{-2}$ for axial and orientational restraints. The conformational space of SBD2 was restrained by keeping a harmonic restraints with a force constant of 5kcal $\cdot$ mol$^{-1}\cdot$Å$^{-2}$ on all of its $C_\alpha$ atoms. The PMF was calculated with the weighted histogram analysis method (WHAM)[103]. The free energy contribution $\Delta G_{a,site}$ was calculated via free energy perturbation and the Zwanzig formula[150]. Hereto, the axial contstraints were released during 8 short simulations(each 10ns,force constants: 50,25,10,5,2.5,1,0.5,0 kcal $\cdot$ mol$^{-1}\cdot$ rad$^{-2}$). In a similar way the $\Delta G_{o,site}$ term was derived. Because of bulk isotropy the orientational correction term in the bulk $\Delta G_{o,bulk}$ can be evaluated analytically. The $\Delta G_{c,site}$ terms were calculated by re-evaluation of PMF's along a RMSD reaction coordinate. Here, we employed Umbrella sampling in a range from 0.5 Å to 5.7 Å in steps of 0.2 Å with a force constant of 15kcal $\cdot$ mol$^{-1}\cdot$ Å$^{-2}$ . The $\Delta G_{c,bulk}$ contributions were obtained in similar fashion.´

### 6.2.4 Replica-Exchange Umbrella Sampling Simulations

Umbrella sampling (US) simulations coupled with Hamiltonian Replica exchange (H-REUS) between neighbouring US intervals were performed to obtain the potentials of mean force for the opening/closing transition of SBD2 with ARG residing in the binding pocket (same procedure for K373A mutation). Here, the ligand had to

be restrained to the binding pocket to prevent its escape in the intermediate umbrella windows. This was done via two harmonic distance restraints between the C atom in the carboxylgroup of ARG and the C atom in ARG333's guanidino group as well as the C atom in ARG's guanidino group and the C atom in ASP267's carboxyl group with a force constant of 1kcal $\cdot$ mol$^{-1}\cdot$ Å$^{-2}$. The centers-of-mass (COM) of backbone atoms in the larger domain D1 (residues 306-308) and backbone atoms in the D2 region (residues 396, 397) served as a reaction coordinate $\xi$ (see Fig. 6.4). A set of 16 US windows biased by harmonic potentials with force constants of 1kcal $\cdot$ mol$^{-1}\cdot$ Å$^{-2}$ were generated with equidistant spacing of 1 Å covering distances of 6 Å up to 21 Å. During the simulations exchanges were attempted every 1ps reaching an acceptance rate of 20-48 %. The total simulation time amounted to 500ns per window. The simulated distributions were analyzed via WHAM, yielding the the corresponding free energy profile. Here, the implementation by Alan Grossfield was used which allows Monte Carlo Bootstrapping analysis [251].

## 6.3 Results

### 6.3.1 Simulation of the Open SBD2 in the Presence of ARG

Extensive unrestrained molecular dynamics (MD) simulations were started from the open SBD2 state (PDB: 4KR5) with six ARG molecules randomly positioned in the vicinity of the protein. Over the course of nearly 6 $\mu s$ of unconstrained simulation, the ARG ligands extensively explored the surface of the open SBD2 conformation (sampled $C_\alpha$ positions of every 100th frame shown in Fig. 6.2, B).

A distinct density peak was identified at the binding site in the D1 subdomain, consistent with the location previously observed for Glutamine [199]. In Fig. 6.2 C, the global density maximum is illustrated on the left side by showing only positions above the threshold of half of the maximum value.

Conversely, other surface-areas were visited infrequently, indicated by low density sampling (additionally, this is reflected in the highly fluctuating RMSD values of the Arg ligands shown in Fig. 6.2 A, upper panel).

Two binding events to the binding pocket, closely resembling the geometry observed in the crystal structure of the closed GLN-liganded form of SBD2 (PDB: 4KQP), were identified (RMSD $<$ 2.5 Å, with respect to the heavy atoms of ligands subsequent

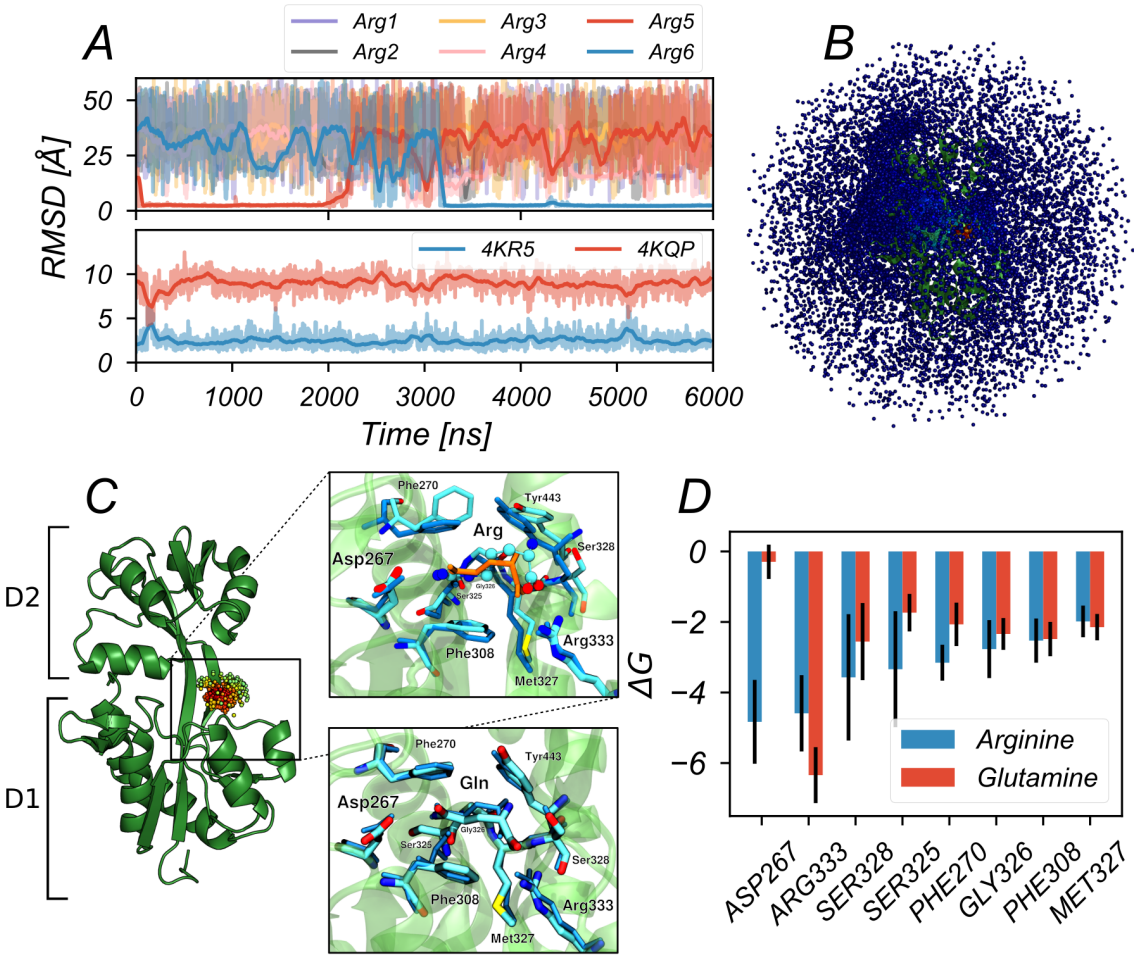



**Fig. 6.2.** Arginine-binding mechanism to the open-protein form of SBD2. (A) Upper panel: Root-mean-square deviation (RMSD, original data and running mean) of each Arg ligand (indicated by different colors) from the native-like binding position of Gln at the D1 domain (PDB: 4KQP). Lower panel: RMSD of the overall SBD2 structure (after alignment of the D1 domains) to the crystal structures of the open (PDB:4KR5) and closed (PDB:4KQP) conformations. Ligand binding does not induce protein closing in the simulated timescales. (B) Illustration of the extensive sampling of the surface of the open SBD2 conformation. The sampled Arg's $C_\alpha$ atom positions are shown for every 100th frame (for better clarity) of the simulation. An increased point density near the ligand binding site of the SBD2 (green cartoon) indicates ARG accumulation at the site. (C) On the left side, the same data as in (B) is presented, with a density threshold set to half of the maximum value. SBD2 is illustrated in a green cartoon representation. On the right side, upper panel, a detailed view of Arg binding to the D1 domain of open SBD2 is shown, aligned with the geometry found in the crystal structure of the closed Gln-bound form of SBD2. (Gln is depicted in orange stick representation, and neighboring residues of the crystal structure are in blue stick representation.) For the sake of easy comparison, a close-up view of the same data obtained from simulations with Gln is shown in the bottom panel, with the crystal structure also in blue stick representation (Data taken from [199]). (D) MM-PBSA interaction energy contributions for both ligands (Gln results have been taken from a previous publication [199]). The charged Arg is capable of forming two stable salt bridges to Asp267 and Arg333.

to alignment of D1's heavy atoms to the crystal structure of the closed -liganded SBD2). Note, that binding to the open SBD2 configuration exclusively took place within the binding pocket located on the larger domain D1 (residues 255-343 and 444-484). These binding events exhibited lifetimes of 2000 ns and >2900 ns, respectively. Notably, these lifetimes are longer than those recorded for GLN (with lifetimes ranging from 100 ns up to 1000 ns, as per previous findings [199]). The Boltzmann inversion of the probabilities of Arg-bound versus unbound states, derived from this simulation, resulted in a standard Arg-binding free energy of $\Delta G \sim$ -3 kcal/mol.

Ligand-binding however did not trigger full domain closing (Fig. 6.2 A, bottom panel), with RMSD > 5 Å to the closedSBD2 (PDB: 4KQP) structure. Several short lived transitions to a semi-open configuration were observed (Fig. 6.2 A, bottom panel at times 150ns, 1450ns, 4100ns and 5140ns. A snapshot of the semi-open state is shown in SI Fig. D.1 B) The lifetimes of this conformation ranged between 20 picoseconds and 80 nanoseconds.

These findings are in qualitative agreement with FRET-measurements, which did not detect any transitions to the closed conformation when Arg was presented to the receptor [217].

Next, we investigated the molecular differences of Arg-binding to SBD2 in comparison to its cognate substrate Gln. Analyzing the energetic contributions of individual side chains to binding in both scenarios allows for the elucidation of the molecular origins of target binding and the extended lifetimes observed in the Arg-SBD2 complexes. For the time intervals with site-specifically bound ligands, MMPBSA trajectory analysis was performed to estimate the mean interaction contributions of individual residues in SBD2, stabilizing both Gln and Arg binding (Fig. 6.2 D).

Both ligands establish stable salt bridges between the terminal carboxyl group and the guanidinium group of Arg333 ($\Delta G$ = -6.34 $\pm$ 0.8 kcal/mol for Gln and $\Delta G$ = -4.6 $\pm$ 1.1 kcal/mol for Arg).

Other residues involved in the binding process exhibit only slight preferences for Arg-binding with $\Delta\Delta G$ = -1.0 kcal/mol for Ser328, $\Delta\Delta G$ = -1.6 kcal/mol for Ser325, $\Delta\Delta G$ = -1.1 kcal/mol for Phe270 and $\Delta\Delta G$ = -0.4 kcal/mol for Gly326. Phe308 and Met327 display negligible distinction between the two ligands with, $\Delta\Delta G$ = -0.04 kcal/mol and $\Delta\Delta G$ = -0.16 kcal/mol, respectively.

However, the critical distinction arises from the charged nature of Arg, enabling the formation of an additional salt bridge with Asp267 in close proximity to the binding site (close-up views for both ligands shown in Fig. 6.2 C on the right side).

This leads to a notably stronger binding energy in the case of Arg, with $\Delta\Delta G$ = -4.5 kcal/mol.

Including all energetic contributions, MMPBSA calculations estimate a slight preference for L-arginine with the following binding energies for both ligands: $\Delta G_{Arg}$ = -10.9 kcal/mol and $\Delta G_{Gln}$ = -9.4 kcal/mol (including a quasi-harmonic entropy term, value for Gln from [199]). However, it's essential to note that this method does not rigorously calculate absolute binding free energies; instead, it computes a mean interaction energy between the ligand and the receptor. Consequently, more robust derivations of equilibrium binding constants based on configurational ensemble averages were carried out in the subsequent section.

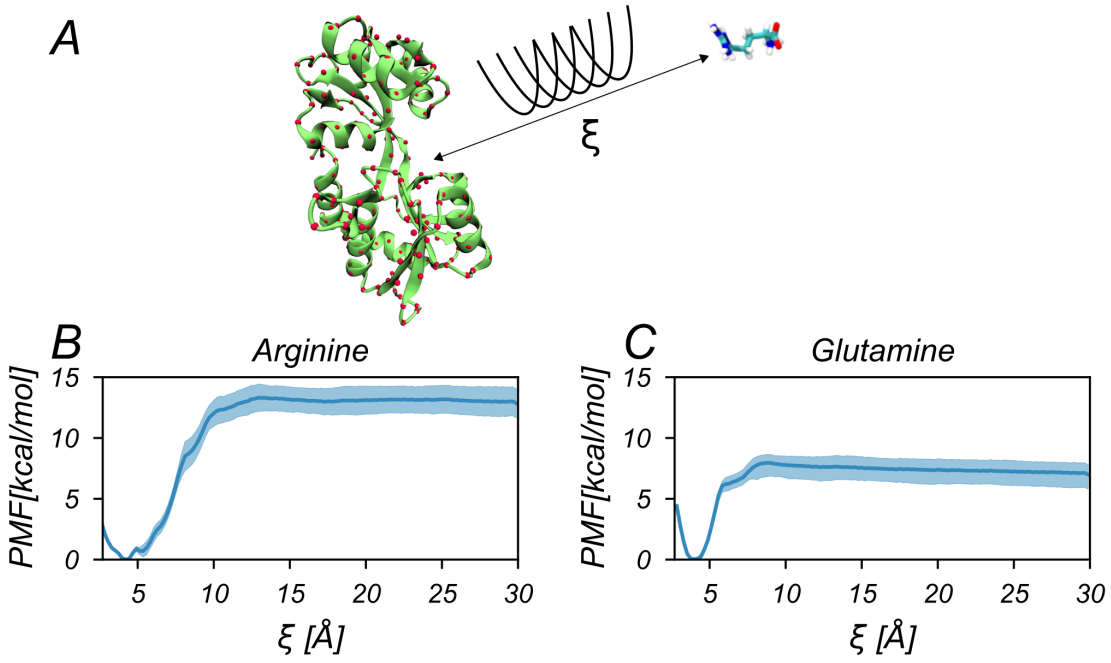
### 6.3.2 Absolute Binding Free Energy Calculations

The quantification of the standard binding free energy $\Delta G_{bind,0}$, including all entropic contributions was performed using a restrained Umbrella sampling approach (details in Section 6.2). Unlike the previous MMPBSA method, this approach explicitly includes solvent, thereby accurately considering both solvent energetic and entropic contributions.

This method allows for a more precise comparison between the two competing ligands, as it enables the robust calculation of absolute binding free energies.

The fundamental concept of this approach involves the incorporation of geometrical and orientational restraints while computing a Potential of Mean Force (PMF) along a radial distance coordinate $\xi$ separating receptor and ligand. This inclusion facilitates rapid convergence. The energetic contributions associated with the introduction or relaxation of these restraints can be individually calculated in the respective endstates (bound vs. unbound). These contributions are then considered and factored in to determine the absolute binding free energy.

The calculated absolute binding free energies amount to $\Delta G_{bind,0}$ = -9.83 kcal/mol ($K_D$ = 0.07 $\mu M$) for Arginine and $\Delta G_{bind,0}$ = -4.96 kcal/mol ($K_D$ = 0.25 mM) for Glutamine (Fig. 6.3). Therefore, the overarching trend of a higher binding affinity of Arg to the open SBD2 form, as obtained through the preceding free simulations and MMPBSA analyses, is further reinforced by an even greater difference between the binding free energies observed here with $\Delta\Delta G_{0,bind}$ = -4.87 kcal/mol. The binding affinities for both ligands are substantially higher than the values derived from a straightforward Boltzmann inversion of the binding probabilities observed




D

| Contribution | Result [kcal/mol] | Standard Error [kcal/mol] |
|---|---|---|
| $\Delta G_{0,bind,restr}$ | -14.90 | 0.10 |
| $\Delta G_{ax,site}$ | 0.97 | 0.13 |
| $\Delta G_{ori,site}$ | 1.51 | 0.13 |
| $\Delta G_{conf,site}$ | 3.24 | 0.19 |
| $\Delta G_{conf,bulk}$ | 4.36 | 0.38 |
| $\Delta G_{ori,bulk}$ | 6.43 | NA |
| $\Delta G_{bind,0}$ | -9.83 | 0.47 |

E

| Contribution | Result [kcal/mol] | Standard Error [kcal/mol] |
|---|---|---|
| $\Delta G_{0,bind,restr}$ | -8.97 | 0.05 |
| $\Delta G_{ax,site}$ | 1.19 | 0.21 |
| $\Delta G_{ori,site}$ | 2.13 | 0.38 |
| $\Delta G_{conf,site}$ | 3.37 | 0.25 |
| $\Delta G_{conf,bulk}$ | 4.42 | 0.39 |
| $\Delta G_{ori,bulk}$ | 6.28 | NA |
| $\Delta G_{bind,0}$ | -4.96 | 0.64 |

**Fig. 6.3.** Absolute binding free energy comparison. (A) Schematic illustration of the simulations employed to calculate the PMF of binding with multiple restraints acting on the system. The $C_\alpha$-atoms (shown as red dots) of SBD2 are restrained to the binding pose by harmonic potentials. The ligands (ARG, GLN) are axially restrained by a polar and an azimuthal angle. The orientation of the ligand is fixed by restraining three Euler angles. These restraining contributions are corrected for to get the physically meaningful standard binding free energy $\Delta G_{0,bind}$. (B,C) Calculated potential of mean force (PMF) along the restrained SBD2-glutamine separation distance $\xi$. The uncertainty is estimated by splitting the data set into 10 subintervals.(D,E) Energetic contributions for applying/releasing orientational, translational and configurational restraints for both ligands.

in the free simulations (see Section 6.3.1 ). This disparity can be attributed to the substantial sampling effort required to attain sufficient statistics when simulating ligand binding to a receptor in an unrestrained manner. This need for extensive sampling was the motivation behind implementing the advanced sampling scheme in the first place. The challenges associated with free simulations, particularly in capturing a significant number of uncorrelated ligand binding events, are well-recognized. As more uncorrelated binding events are sampled in free simulations, a convergence of the values obtained through Boltzmann inversion towards those obtained by the advanced sampling approach is expected.

Experimental measurements via Isothermal Titration Calorimetry (ITC) indicate a dissociation constant of $K_d$ = 0.9 $\pm$ 0.2 $\mu M$ for Glutamine, corresponding to a $\Delta G_{bind,0}$ = -8.3 kcal/mol [218]. It is important to note that the calculations presented here exclusively cover the binding of ligands to the open form of SBD2. Given that experimental approaches typically incorporate contributions from domain closure, discrepancies in binding affinity are expected.

The apparent binding affinity of L-arginine appears to be exaggerated in our simulations. Experimental findings indicate that higher concentrations of arginine hinder the uptake of glutamine into the cell, with a $K_i$ value in the millimolar range [213, 217]. The current calculations appear to overestimate the ARG binding free energy. The binding of charged side chains is influenced by a balance between charge-charge interaction and changes in solvation. Due to the very favourable solvation free energy of the guanidinium group of the ARG side chain ($\approx -80 kcal/mol$ [252, 253]), this balance may not be accurately represented, resulting in an overestimation of electrostatic attraction versus (de-)solvation. Indeed, previous studies have demonstrated an overestimation of electric field strengths within the utilized ff14SB force field [254], which may overstabilize salt-bridge formation. We anticipate that this overestimation effect has a more pronounced impact in the case of the substrate ARG due to its charged nature.

Stronger binding of Arg largely originates from deviations in the $G_{0,bind,restr}$ terms ($\Delta\Delta G_{0,bind,restr}$ = -5.93 kcal/mol). This term is derived from the PMF shown in Figure 6.3 B,C and corresponds to the interactions gained when the ligands transition from the bulk solution into the binding pocket. Notably, this observation corresponds well with the earlier analyses that highlighted the favorable electrostatic interactions of Arg within the D1 binding site of the open SBD2. The total cost of free energy associated with the loss of conformational freedom $\Delta G_{conf,bulk} - \Delta G_{conf,site}$ amounts to 1.12 $\pm$ 0.42 kcal/mol for Arg and 1.05 $\pm$ 0.46

kcal/mol for Gln. This indicates, that the restriction of the conformational flexibility of SBD2 upon ligand binding plays an negligible role in discriminating between Arg and Gln.

The overall loss of orientational freedom upon binding $\Delta G_{ori,bulk} - \Delta G_{ori,site}$ leads to a free energy cost of 4.92 $\pm$ 0.13 kcal/mol in the case of Arg and 4.15 $\pm$ 0.38 kcal/mol for Gln. That means that the orientational fluctuations of Arg within the binding pocket are more contrained compared to those of Gln. From a physical point of view this is reasonable since both ends of Arg are stabilized by stable salt bridges to Arg333 and Asp267 (see Fig. 6.2 C, upper panel on the right side).

### 6.3.3 Arg Blocking the Closed SBD2 Conformation

Single molecule FRET experiments unveiled that transitions from the open SBD2 to the closed conformation of the protein are largely suppressed in the presence of Arg. Despite the unrestrained simulations demonstrating strong binding of Arg to the open SBD2 conformation, there were no observed transitions to the closed state within the microsecond timescale (see Fig. 6.2 A, bottom panel). This suggests the existence of a substantial transition barrier for the opening/closing of the protein. In an attempt to capture the transition during free simulations, we modified SBD2 to SBD2(L480A). This specific point mutation has been proven to notably decrease the transition barrier, as evidenced in both computer simulations [199] and FRET experiments [229]. However, while the L480A mutation for Gln was adequate to capture transitions to the closed state, the introduction of this mutation with Arg bound did not enable the sampling of full transitions to the closed SBD2 state. The mutation however led to a more pronounced sampling of the before mentioned semi-open state (see Fig. SI D.3). This indicates, in addition to a high transition barrier, a considerable energetic penalty on the closed state when arginine is bound to the open conformation.

Next, extensive Hamiltonian-replica-exchange Umbrella Sampling (HREUS) simulations were conducted on the wildtype SBD2 protein. These simulations utilized a reaction coordinate $\xi$ based on a center-of-mass distance between the two subdomains D1 and D2, capable of a clear distinction between the open and closed state (Fig. 6.4 B). Through the application of harmonic biasing potentials, SBD2 was steered towards the closed state. The resulting Potential of Mean Force (PMF) along $\xi$ illustrates the free energy expenses required for this manipulation. Note,

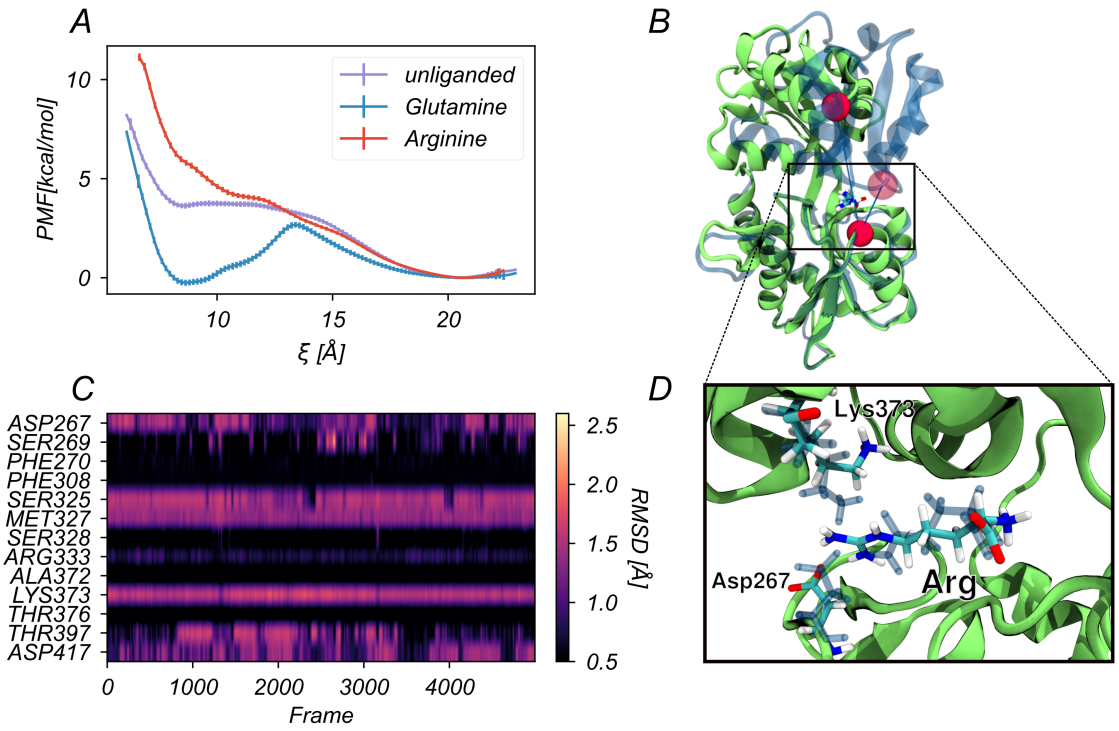



**Fig. 6.4.** Mechanism of arginine blocking SBD2 closing. (A) Influence of different ligands on the potentials of mean force (PMF) upon closing. While glutamine (blue line) stabilizes the bound state compared to the unliganded SBD2 (in purple), arginine (red line) hampers closing with a steep free energy increase for $\xi$ distances below 10 $\AA$. Note, that arginine had to be restrained to the binding pocket during Umbrella sampling simulations (B) Cartoon representation of the simulation setup. The open form SBD2 (PDB: 4KR5, in green) with arginine (stick representation) bound to D1. Two centers of mass (red spheres) were defined to pull SBD2 into the closed state (blue cartoon representation) via Umbrella Sampling simulations. (C) Local side chain deformations compared to the crystal structure of closed SBD2 (PDB: 4KQP). RMSD of residue side chains to the crystal structure in the close proximity of arginine after best fit to the backbone. Lys373 constantly samples high deviation configurations. (D) Close up view a frame sampled in the closed SBD2-arginine bound state. The positively charged group of Lys373 cannot adopt the crystal structure (in blue transparent sticks) conformation because of repelling Coulomb interactions with the also positively charged arginine.

that during these simulations ARG had to be restrained to the D1 binding site (details in Section 6.2).

The calculated PMF shows the considerable free energy cost of closing SBD2-Arg with ΔG values exceeding 6 kcal/mol (Fig. 6.4 A, in red). Previous MD simulations in contrast have shown a slight preference of the closed state for the SBD2-Gln system with $\Delta G \sim$ -0.3 kcal/mol (Fig. 6.4 A, in blue). The unfavorable and high-free-energy conformations from these pulling simulations, where Arg occupies the binding pocket, were scrutinized for local deformations in comparison to the crystal structure of the closed Gln-bound SBD2 complex. In this analysis, the backbone of residues in close proximity to the ligand (within $<$ 4 Å) were individually superimposed onto the crystal reference (PDP: 4KQP), and the RMSD for the side chains was calculated for each frame (Fig. 6.4 C).

Among these, SER325, MET327, and LYS373 displayed the most significant deviations (mean RMSDs: 1.45 Å for SER325, 1.38 Å for MET327, 1.58 Å for LYS373). Notably, only LYS373 consistently exhibited substantial deviations from the crystal structure (minimal RMDS: 0.1 Å for SER325, 0.15 Å for MET327, 0.7 Å for LYS373).

Upon inspection of the adopted configurations, a repulsion of the positively charged $\epsilon$-amino group of LYS373 from the Arg amino group, which also carries a positive charge, throughout all frames can be observed. Nevertheless, the pronounced deviations and fluctuations in the conformation of the other residues (Ser325 and Met327) also indicate an elevated degree of steric repulsion resulting from the presence of ARG, which impedes domain closure. In Fig. 6.4 D, an exemplary conformation is depicted where the side-chain amino group of LYS373 is significantly tilted away from Arg, illustrating the source of the substantial local deviation.

### 6.3.4 K373A Allows Closing

To evaluate the hypothesized significant role of K373 in impeding the closure of SBD2 when arginine is bound to D1, the impact of a K373A mutation on the overall dynamic behavior of the open SBD2-Arg complex was assessed. First, extensive unrestrained simulations of the SBD2(L480A/K373A)-Arg bound complex were performed. Similar to the preceding simulation, where closure of SBD(L490A) was prohibited, the mutation of the C-terminal L480A point mutation was included to decrease the transition times between the open and closed states.

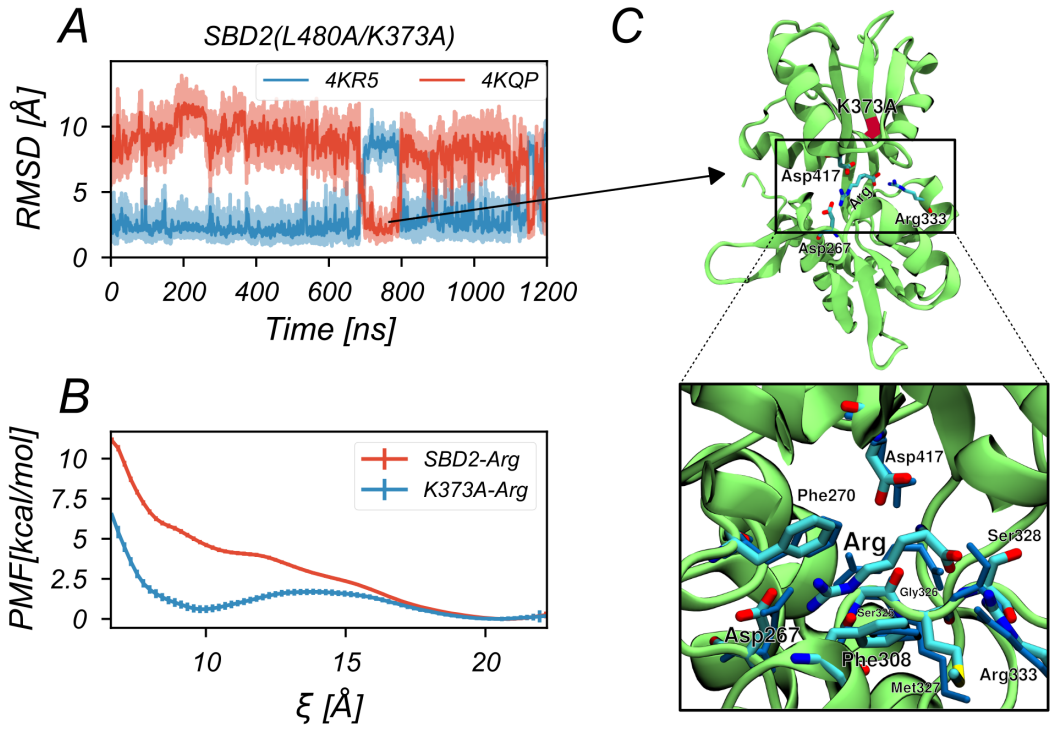



**Fig. 6.5.** Point mutation K373A in SBD2(L480A) drastically lowers the free energy of the closed SBD2 configuration. A) RMSD to the crystal structures of the closed (PDB:4KQP, in red) and the open (PDB: 4KR5, in blue) conformation. B) PMF calculated for SBD2 K373A. Note, that Leu480 was not mutated to Ala in these simulations. C) Snapshots of a exemplary closed conformation. Binding site residues have been aligned to the crystal structure (in blue stick representation).

In the course of 1.2 $\mu s$ of free simulation, several short transitions to the semi-open conformation were observed at times 95ns, 274ns, 533ns, 877ns and 1196ns. Interestingly, a complete closure of SBD2(L480A/K373A) in the presence of Arg was captured, indicated by an RMSD of less than 1.8 Å to the closed crystal reference 4KQP (see Fig. 6.5 A). The closed state was maintained for a duration of 88 ns. Subsequently, a brief adaptation to the closed state occurred between 1151 ns and 1169 ns, with slightly increased deviations from the closed conformation (RMSD $\sim$ 2 Å). However, this configuration quickly transitioned back to the open state, resulting in a lifetime of 18 ns.

Closer inspection of the binding site of the fully closed SBD2(L480A/K373A) complex showed minimal structural deviations from the crystallized structure with bound Gln. Figure 6.5, C depicts an exemplary frame from the simulation in superposition to the spatial sidechain arrangement in the crystal structure (in blue stick representation).

Next, the PMF for the opening/closing transition of SBD2 K373A was calculated following the same procedure as in the unmutated case. Figure 6.5, B illustrates the impact of the K373A mutations on the free energy profile. The mutation drastically decreases the free energy of the closed state by $\Delta G$ = 4.58 kcal/mol. The resulting PMF minimum aligns with the stable closing observed in the free simulations.

However, despite this mutation, the open SBD2(L480A/K373A) conformation remains slightly energetically favored. The energetic cost associated with closure amounts to $\Delta G$ = 0.48 kcal/mol.

---

## 6.4 Discussion

Substrate-binding proteins (SBPs) play a vital role in facilitating the cellular uptake of essential nutrients [209, 211, 214, 255], yet the intricate relationship between ligand binding mechanisms and subsequent conformational changes within SBPs remains elusive for many. The conformational states of these proteins hold central importance in governing the overall transportation process [203, 256], raising intriguing questions that demand exploration before achieving a comprehensive understanding, e.g.:
Is a conformational transition a prerequisite for substrate translocation? Unraveling this aspect can provide insights into the regulatory role of protein dynamics in

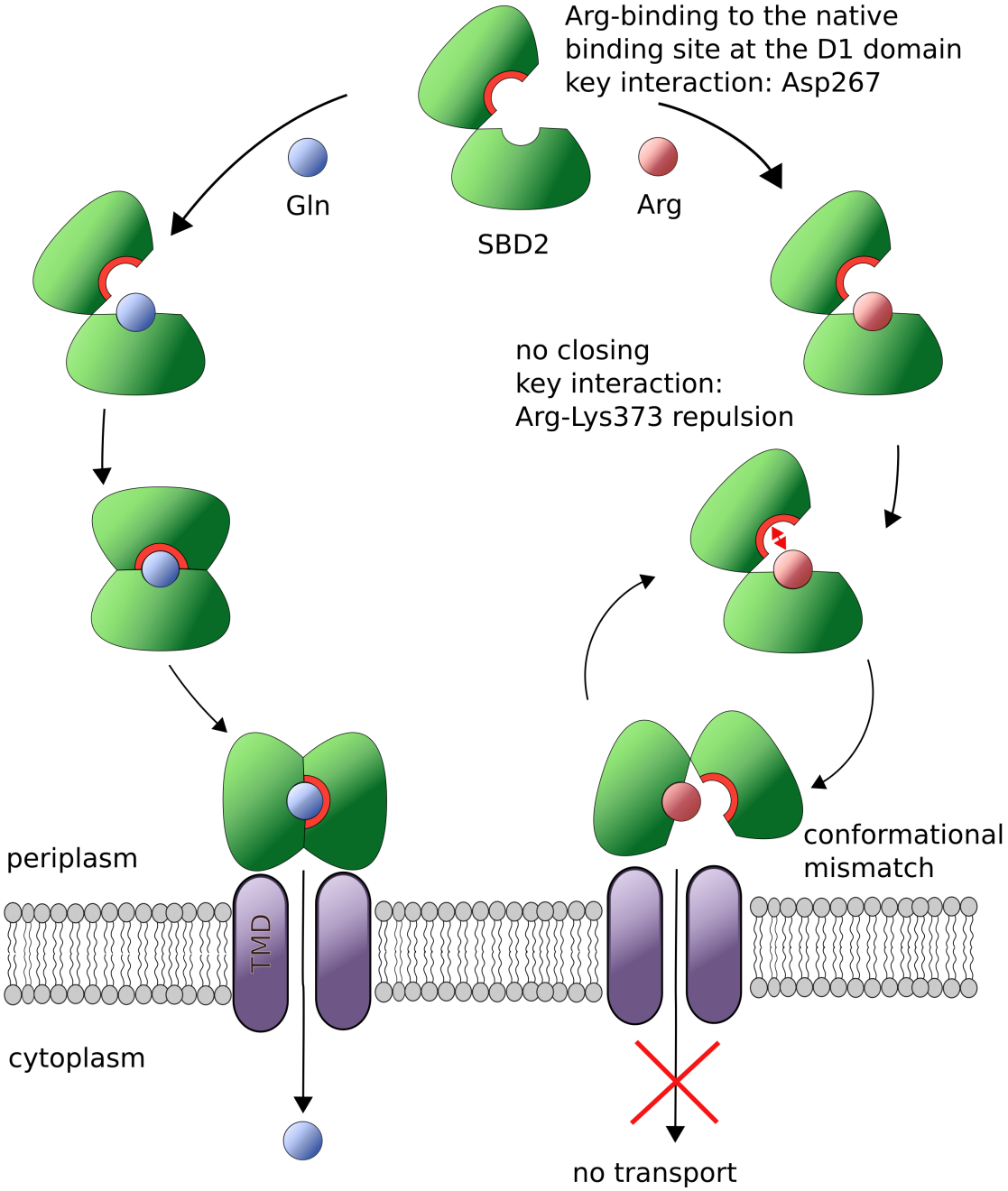



**Fig. 6.6.** Proposed mechanism of L-arginine blocking Gln-transport. Asp267 plays a pivotal role in conferring binding ability to the substrate arginine. Subsequent transition to the closed state of SBD2 is prevented when arginine is bound. Notably, sterical repulsion between the ARG and residues at the D1-D2 interface including charge repulsion between bound ARG and Lys373 emerged as a key interactions that obstructs the conformational change. A conformational mismatch between the openSBD2-Arg complex and the TMDs may prevent successful passage into the cytoplasm.

the transport process. How does the binding of ligands relate to conformational transitions? Clarifying this intricate interplay could hold significant importance in understanding substrate specificity. What are the time scales of the various processes preceding the efficient translocation of nutrients into the cell?

Our study focuses on addressing this complex issue, particularly in the context of the Substrate Binding Domain 2 (SBD2) of the GlnPQ importer from *Lactococcus lactis* , shedding light on its role in glutamine (Gln) uptake and the inhibitory effects of L-arginine. Single-molecule Förster resonance energy transfer (smFRET) measurements provided key insights into the competitive interaction between Gln and L-arginine [217]. These measurements revealed that L-arginine not only competitively inhibits the Gln-triggered closed conformation of SBD2 but also impedes Gln translocation into the cell, with an inhibition constant $K_i$ in the millimolar range [213, 217, 257].

To explore this phenomenon in molecular detail, we conducted extensive molecular dynamics (MD) simulations alongside advanced sampling methods.

Our MD simulations showed that L-arginine occupies the same binding site at the D1 domain observed in the previous chapter for its cognate substrate, Gln. We calculated the binding affinities to SBD2 for both Gln and Arg. The computed binding free energy of Gln of $\sim$ -5 kcal/mol is in qualitative agreement with an experimental dissociation constant, translating to a binding free energy of approximately -8 kcal/mol. It's worth noting that our calculations only consider binding to the open configuration of SBD2, and we anticipate an additional gain in free energy upon domain closure of SBD2.

The calculated binding affinity of ARG to D1 in the open state was found to be higher than that of GLN in both free simulations, using MMPBSA and when employing more sophisticated absolute binding free energy simulations. A stable salt bridge was identified as the pivotal interaction responsible for the enhanced binding affinity, with the positively charged ARG forming a stable bond with Asp267. Nevertheless, experimental evidence indicates a significantly lower binding affinity for ARG, although a direct experimental measurement is not available. Experimental data was obtained from the ARG concentrations necessary to inhibit GLN transport, and may involve other contributions than just competitive binding. However, from a biological perspective, it is unlikely that ARG would exhibit a stronger binding affinity, as this would prevent GLN transport. We attribute this discrepancy to a possible overestimation of charge-charge interactions in salt-bridge formation, a common issue associated with empirical additive force fields [258–262]. In particular, attractive solute-solute interactions are notorious for

yielding overestimated results[258, 263]. Indeed, the employed ff14SB force field has exhibited a possible overestimation of such electrostatics interactions, especially within binding sites containing arginine [254]. We speculate that these artefacts may have a minor impact on the uncharged glutamine. Nevertheless, our results provide qualitative evidence for ARG binding at the same site as GLN and offer an explanation for the inhibition of GLN binding by ARG. Furthermore, our studies have revealed that when ARG binds to the native binding site at the D1 domain, a transition to the closed SBD2 state is strongly suppressed. Using umbrella sampling simulations combined with Hamiltonian replica exchange (H-REUS) resulted in a substantial penalty, exceeding 6 kcal/mol, for the transition to the closed state under these conditions. The presence of a positively charged Lys residue at position 373 in the D2 subdomain was identified as a primary cause, resulting in electrostatic repulsion. However, steric repulsion with respect to K373 and neighbouring residues may also play a role. To ascertain the significance of K373, we conducted simulations utilising the SBD2-K373A variant, which revealed a considerable impact on the free energy profile of closure, with only a slight preference for the open state ($\sim$ 0.48 kcal/mol) for the mutation (in the presence of ARG).

In conclusion, the results of the simulations indicate that the inhibition of GLN by ARG occurs via the mechanism illustrated in Fig. 6.6. When present in sufficient concentration, ARG inhibits GLN-binding by occupying the same native SBD2 binding site in the D1 subdomain, preventing the spontaneous transition to the closed state. Furthermore, we identified residues in SBD2 that are perturbed by ARG binding and contribute to preventing closure. The introduction of an Ala mutation at position 373 (L373A) significantly reduced this effect, facilitating the adoption of the closed state in computer simulations.

Consequently, subsequent substrate transport may be impeded, as the openSBD2-ARG complex is unable to initiate allosteric interactions with the translocator. Docking of the closed conformation to the transmembrane domains of the importer is an essential step for substrate delivery, as previously demonstrated in the context of SBD2 [220]. The "conformational mismatch" hypothesis has been previously proposed for SBD2 and other substrate-binding domains [216, 264, 265].

Nevertheless, the results of our simulations indicate that the SBD2(K373A) variant may enable L-arginine uptake. The mutation represents a promising candidate for experimental validation of this hypothesis, which aims to confirm whether a conformational mismatch between open SBD2 and the translocator indeed explains the observed inhibition of arginine uptake into the cell. The work contributes to our

understanding of the molecular mechanism and dynamics that govern SBD2. It also highlights the importance of considering conformational changes and associated free energy landscapes in elucidating the functional aspects of substrate-binding proteins.

# 7

# Summary and Outlook

Starting with Rahman's pioneering work in 1964 [266], molecular simulation of biomolecules gained widespread acceptance in the fields of chemistry and biophysics, eventually culminating in the 2013 Nobel Prize awarded to Karplus, Levitt, and Warshel. Today, molecular dynamics (MD) simulations serve as a critical scientific tool, providing unprecedented insight into biomolecular processes at spatial and temporal resolutions that are in many cases inaccessible to experimentation alone [267–271].

The fruitful collaboration between molecular simulation and experimentation has led to significant advances in the interpretation of many experimental results. In addition, computational techniques are becoming increasingly adept at accurately predicting experimental results. Chapter 5 and Chapter 6 exemplifies this synergy by showing how MD simulations elucidate the ligand binding process for the substrate-binding domain 2 of GlnPQ and subsequent domain motions at the atomistic level. In particular, the simulations predicted key residues for global transitions, which were later confirmed by FRET experiments. Understanding the crucial residues involved in binding and subsequent conformational transitions within this protein can serve to guide more future experimental work, thereby illuminating the intricate mechanism underlying cellular import facilitated by the ABC-importer GlnPQ.

However, there are still major challenges in MD simulations, firstly due to the huge configuration space of biomolecules, which may lead to sampling problems, and secondly due to the accuracy of the underlying molecular models or force fields [272, 273]. However, ongoing developments promise to push the boundaries of simulation capabilities. With regard to the latter problem, for example, novel machine learning approaches offer the possibility of constructing high-precision force fields independent of traditional functional forms, which even allow the modelling of chemical reactivity in MD simulations [274–277]. Of course, future breakthroughs in the active field of quantum computing could usher in a new era of unprecedented accuracy in simulation-based research [278–280].

Meanwhile, advances in computational power and sampling methods in recent decades have mitigated the sampling problem faced by MD simulations. Chapter 3 presents a tailored HREMD sampling method, $\omega$BP-REMD, which effectively samples the cis/trans isomerization of proline. By adjusting the barrier heights of the torsional force field terms surrounding the $\omega$-dihedral angle of the peptide bond preceding a proline along the replica ladder and incorporating a one-fold bias potential for balanced sampling between the two isomeric states, we achieved excellent cis/trans sampling in all systems tested. Our results showed superior agreement with experimental data and faster convergence compared to the standard Umbrella sampling approach, which was particularly evident when applied to a proline within a hairpin loop of the N2 domain of the G3P protein.

In my opinion, an important practical advantage of this method is its ease of use. Unlike other advanced MD simulation techniques, where numerous parameters can complicate setup and troubleshooting, $\omega$BP-REMD offers a streamlined approach. Once the number of replicas $n$ and corresponding barrier heights are appropriately set to ensure sufficient exchange rates between replicas, along with the optimal height of the one-fold bias term to achieve balanced cis/trans sampling, the method runs smoothly. This simplicity suggests significant potential for its application to smaller peptides and intrinsically disordered proteins, where variations in these parameters are expected to be minimal. The values used in this work can serve as valuable guidelines for such systems.

For systems where the potential of mean force (PMF) around the $\omega$-dihedral is not primarily governed by Fourier series dihedral terms, establishing initial reference values for these parameters is not straightforward. Future developments of the method may include on-the-fly estimations for optimal parameter adjustment.

In Chapter 4, we applied the $\omega$BP-REMD method to three proline residues within the nonapeptide bradykinin, revealing the interdependence of individual Pro isomerization states. Our analysis showed that different combinations of Pro cis/trans states populate distinct conformational clusters, suggesting coupling between Pro isomerization states even in the absence of structural restraints as seen in folded proteins. This investigation highlights a major strength of $\omega$BP-REMD in probing such potential couplings. In particular, this method biases all $\omega$-dihedral reaction coordinates independently, requiring significantly less sampling effort than standard metadynamics or Umbrella sampling methods.

Given that proline residues are commonly found in intrinsically disordered proteins and shorter peptides, there is a wide range of systems ideally suited for investigation using $\omega$BP-REMD [281–286]. I am optimistic that this method can

facilitate a more comprehensive exploration of the entire configurational space of such molecules, encompassing all contributing Pro cis/trans combinations in future computer simulations.

# Bibliography


[1]Charles Tanford and Jacqueline Reynolds. *Nature's robots: a history of proteins*. Oxford Univ. Press, 2003.

[2]Bruce Alberts et al. *Molecular biology of the cell*. Vol. 3. Garland New York, 1994.

[3]K Wang, J McClure, and A Tu. "Titin: major myofibrillar components of striated muscle." In: *Proceedings of the National Academy of Sciences* 76.8 (1979), pp. 3698–3702.

[4]Christopher F Higgins. "ABC transporters: from microorganisms to man". In: *Annual review of cell biology* 8.1 (1992), pp. 67–113.

[5]Kaspar Hollenstein, Roger JP Dawson, and Kaspar P Locher. "Structure and mechanism of ABC transporter proteins". In: *Current Opinion in Structural Biology* 17.4 (Aug. 1, 2007), pp. 412–418.

[6]Harvey F Lodish. *Molecular cell biology*. Macmillan, 2008.

[7]Pratul K. Agarwal. "Enzymes: An integrated view of structure, dynamics and function". In: *Microb Cell Fact* 5.1 (2006), p. 2.

[8]John T. Hancock. *Cell signalling*. Oxford University Press, 2017.

[9]Peter E. Wright and H. Jane Dyson. "Intrinsically disordered proteins in cellular signalling and regulation". In: *Nat Rev Mol Cell Biol* 16.1 (2015), pp. 18–29.

[10]Richard L. Zschau and Martin Zacharias. "Mechanism of $\beta$ -hairpin formation in AzoChignolin and Chignolin". In: *J Comput Chem* 44.9 (2023), pp. 988–1001.

[11]Matthew C. Good, Jesse G. Zalatan, and Wendell A. Lim. "Scaffold Proteins: Hubs for Controlling the Flow of Cellular Information". In: *Science* 332.6030 (2011), pp. 680–686.

[12]Olaf Fritsche and Martin Lay. *Biologie für Einsteiger: Prinzipien des Lebens verstehen*. Springer-Spektrum, 2015.

[13]José Nelson Onuchic and Peter G Wolynes. "Theory of protein folding". In: *Current Opinion in Structural Biology* 14.1 (2004), pp. 70–75.

[14]Ken A. Dill and Justin L. MacCallum. "The Protein-Folding Problem, 50 Years On". In: *Science* 338.6110 (2012), pp. 1042–1046.

[15]Cyrus Levinthal. “How to fold graciously”. In: *Mossbauer spectroscopy in biological systems* 67 (1969), pp. 22–24.

[16]Amy H. Andreotti. “Native State Proline Isomerization: An Intrinsic Molecular Switch”. In: *Biochemistry* 42.32 (2003), pp. 9515–9524.

[17]William J Wedemeyer, Ervin Welker, and Harold A Scheraga. “Proline cis- trans isomerization and protein folding”. In: *Biochemistry* 41.50 (2002), pp. 14637–14644.

[18]Paramita Sarkar et al. “Proline cis-trans Isomerization Controls Autoinhibition of a Signaling Protein”. In: *Molecular Cell* 25.3 (2007), pp. 413–426.

[19]Andrew D. Robertson and Kenneth P. Murphy. “Protein Structure and the Energetics of Protein Stability”. In: *Chem. Rev.* 97.5 (1997), pp. 1251–1268.

[20]Katherine Henzler-Wildman and Dorothee Kern. “Dynamic personalities of proteins”. In: *Nature* 450.7172 (2007), pp. 964–972.

[21]H. Jane Dyson and Peter E. Wright. “Intrinsically unstructured proteins and their functions”. In: *Nat Rev Mol Cell Biol* 6.3 (2005), pp. 197–208.

[22]J. Andrew McCammon, Bruce R. Gelin, and Martin Karplus. “Dynamics of folded proteins”. In: *Nature* 267.5612 (1977), pp. 585–590.

[23]Christoph R Jacob and Johannes Neugebauer. “Subsystem density-functional theory”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 4.4 (2014), pp. 325–362.

[24]David A. Case et al. *Amber 2021*. 2021.

[25]Pradyumn Sharma, Rajat Desikan, and K. Ganapathy Ayappa. “Evaluating Coarse-Grained MARTINI Force-Fields for Capturing the Ripple Phase of Lipid Membranes”. In: *J. Phys. Chem. B* 125.24 (2021), pp. 6587–6599.

[26]Juan R Perilla et al. “Molecular dynamics simulations of large macromolecular complexes”. In: *Current Opinion in Structural Biology* 31 (2015), pp. 64–74.

[27]David E Shaw et al. “Millisecond-scale molecular dynamics simulations on Anton”. In: *Proceedings of the conference on high performance computing networking, storage and analysis*, pp. 1–11.

[28]Manuel Luitz et al. “Exploring biomolecular dynamics and interactions using advanced sampling methods”. In: *J. Phys.: Condens. Matter* 27.32 (2015), p. 323101.

[29]Ron O. Dror et al. “Biomolecular Simulation: A Computational Microscope for Molecular Biology”. In: *Annu. Rev. Biophys.* 41.1 (2012), pp. 429–452.

[30]L. D. Landau and E. M. Lifshitz. *Mechanics*. Vol. 1. CUP Archive, 1960.

[31]Matthias Bartelmann et al. *Theoretische Physik*. Springer Spektrum, 2014.

[32]R. K. Pathria and Paul D. Beale. *Statistical mechanics*. Elsevier/Academic Press, 2011.

[33]Leonard Susskind and George Hrabovsky. *Classical mechanics: the theoretical minimum*. Penguin Books, 2014.

[34]Mark E Tuckerman. *Statistical mechanics: theory and molecular simulation*. Oxford university press, 2023.

[35]Michael P. Allen and Dominic J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 2017.

[36]Franz Schwabl. *Statistical mechanics*. Springer Science & Business Media, 2006.

[37]E. T. Jaynes. "The Gibbs Paradox". In: *Maximum Entropy and Bayesian Methods: Seattle, 1991*. 1992, pp. 1–21.

[38]Kerson Huang. *Statistical mechanics*. Wiley, 1987.

[39]Dominik Marx and Jürg Hutter. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press, 2009.

[40]D. C. Rapaport. *The Art of Molecular Dynamics Simulation*. Cambridge University Press, 2004.

[41]Chris Lorenz and Nikos L Doltsinis. "Molecular Dynamics Simulation: From "Ab Initio" to "Coarse Grained"". In: *Handbook of Computational Chemistry*. 2012, pp. 195–238.

[42]Sam P. de Visser et al. "Multi-State Epoxidation of Ethene by Cytochrome P450: A Quantum Chemical Study". In: *J. Am. Chem. Soc.* 123.13 (2001), pp. 3037–3047.

[43]Oliver Weingart, Igor Schapiro, and Volker Buss. "Photochemistry of Visual Pigment Chromophore Models by Ab Initio Molecular Dynamics". In: *J. Phys. Chem. B* 111.14 (2007), pp. 3782–3788.

[44]Marie-Pierre Gaigeot and Michiel Sprik. "Ab Initio Molecular Dynamics Computation of the Infrared Spectrum of Aqueous Uracil". In: *J. Phys. Chem. B* 107.38 (2003), pp. 10344–10358.

[45]C. Molteni, I. Frank, and M. Parrinello. "An Excited State Density Functional Theory Study of the Rhodopsin Chromophore". In: *J. Am. Chem. Soc.* 121.51 (1999), pp. 12177–12183.

[46]Carme Rovira and Michele Parrinello. "Oxygen binding to iron–porphyrin: A density functional study using both LSD and LSD+GC schemes". In: *International Journal of Quantum Chemistry* 70.2 (1998), pp. 387–394.

[47]Carme Rovira et al. "Equilibrium Geometries and Electronic Structure of Iron Porphyrin Complexes: A Density Functional Study". In: *J. Phys. Chem. A* 101.47 (1997), pp. 8914–8925.

[48]Angelo Bifone, H. J. M. de Groot, and Francesco Buda. “Ab initio molecular dynamics of rhodopsin”. In: *Pure and Applied Chemistry* 69.10 (1997), pp. 2105–2110.

[49]M. Born and W. Heisenberg. “Zur Quantentheorie der Molekeln”. In: *Original Scientific Papers Wissenschaftliche Originalarbeiten*. Springer Berlin Heidelberg, 1985.

[50]Dariusz Ke Dziera and Anna Kaczmarek-Kedziera. “Remarks on Wave Function Theory and Methods”. In: *Handbook of Computational Chemistry*. 2017, pp. 123–171.

[51]Luca Monticelli and D. Peter Tieleman. “Force Fields for Classical Molecular Dynamics”. In: *Biomolecular Simulations*. Vol. 924. 2013, pp. 197–213.

[52]Linda E. Reichl. *A Modern Course in Statistical Physics*. John Wiley & Sons, 2016.

[53]Kresten Lindorff-Larsen et al. “How Fast-Folding Proteins Fold”. In: *Science* 334.6055 (2011), pp. 517–520.

[54]Till Siebenmorgen and Martin Zacharias. “Evaluation of Predicted Protein–Protein Complexes by Binding Free Energy Simulations”. In: *J. Chem. Theory Comput.* 15.3 (2019), pp. 2071–2086.

[55]Martin Karplus and J. Andrew McCammon. “Molecular dynamics simulations of biomolecules”. In: *Nat Struct Mol Biol* 9.9 (2002), pp. 646–652.

[56]Juan R. Perilla and Klaus Schulten. “Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations”. In: *Nat Commun* 8.1 (2017), p. 15959.

[57]Julian Hartmann and Martin Zacharias. “Analysis of amyloidogenic transthyretin mutations using continuum solvent free energy calculations”. In: *Proteins: Structure, Function, and Bioinformatics* 90.12 (2022), pp. 2080–2090.

[58]Po-chia Chen et al. “Structure-based screening of binding affinities via small-angle X-ray scattering”. In: *IUCrJ* 7.4 (2020), pp. 644–655.

[59]Korbinian Liebl and Martin Zacharias. “Tumuc1: A New Accurate DNA Force Field Consistent with High-Level Quantum Chemistry”. In: *J. Chem. Theory Comput.* 17.11 (2021), pp. 7096–7105.

[60]Korbinian Liebl and Martin Zacharias. “The development of nucleic acids force fields: From an unchallenged past to a competitive future”. In: *Biophysical Journal* 122.14 (2023), pp. 2841–2851.

[61]Oliver T. Unke et al. “Machine Learning Force Fields”. In: *Chem. Rev.* 121.16 (2021), pp. 10142–10186.

[62]Wendy D. Cornell et al. “A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules”. In: *J. Am. Chem. Soc.* 117.19 (1995), pp. 5179–5197.

[63]Samuel Krimm and Jagdeesh Bandekar. “Vibrational Spectroscopy and Conformation of Peptides, Polypeptides, and Proteins”. In: *Advances in Protein Chemistry*. Vol. 38. 1986, pp. 181–364.

[64]Lionel Goodman, Vojislava Pophristic, and Frank Weinhold. “Origin of Methyl Internal Rotation Barriers”. In: *Acc. Chem. Res.* 32.12 (1999), pp. 983–993.

[65]Tamar Schlick. *Molecular Modeling and Simulation*. Springer New York, 2002.

[66]J. E. Jones and Sydney Chapman. “On the determination of molecular fields.—I. From the variation of the viscosity of a gas with temperature”. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 106.738 (1997), pp. 441–462.

[67]W. Pauli. “Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren”. In: *Z. Physik* 31.1 (1925), pp. 765–783.

[68]F. London. “Zur Theorie und Systematik der Molekularkr fte”. In: *Z. Physik* 63.3 (1930), pp. 245–279.

[69]Judith A. Harrison et al. “Review of force fields and intermolecular potentials used in atomistic computational materials research”. In: *Applied Physics Reviews* 5.3 (2018), p. 31104.

[70]Alexander D. Mackerell Jr. “Empirical force fields for biological macromolecules: Overview and issues”. In: *Journal of Computational Chemistry* 25.13 (2004), pp. 1584–1604.

[71]Olgun Guvench and Alexander D. MacKerell. “Comparison of Protein Force Fields for Molecular Dynamics Simulations”. In: *Molecular Modeling of Proteins*. 2008, pp. 63–88.

[72]Jens Kleinjung and Franca Fraternali. “Design and application of implicit solvent models in biomolecular simulations”. In: *Current Opinion in Structural Biology* 25 (2014), pp. 126–134.

[73]Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*. Academic Press, 2002.

[74]Peter J. Steinbach and Bernard R. Brooks. “New spherical-cutoff methods for long-range forces in macromolecular simulation”. In: *Journal of Computational Chemistry* 15.7 (1994), pp. 667–683.

[75]Brad A. Wells and Alan L. Chaffee. “Ewald Summation for Molecular Simulations”. In: *J. Chem. Theory Comput.* 11.8 (2015), pp. 3684–3695.

[76]V. Ballenegger. “Communication: On the origin of the surface term in the Ewald formula”. In: *The Journal of Chemical Physics* 140.16 (2014), p. 161102.

[77]Tom Darden, Darrin York, and Lee Pedersen. “Particle mesh Ewald: An N log (N) method for Ewald sums in large systems”. In: *The Journal of chemical physics* 98.12 (1993), pp. 10089–10092.

[78]Loup Verlet. “Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules”. In: *Phys. Rev.* 159.1 (1967), pp. 98–103.

[79]R. W. Hockney. *Computer Simulation Using Particles*. CRC Press, 2021.

[80]Ernst Hairer, Gerhard Wanner, and Christian Lubich. “Numerical Integrators”. In: *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. 2006, pp. 27–50.

[81]Benedict J. Leimkuhler, Sebastian Reich, and Robert D. Skeel. “Integration Methods for Molecular Dynamics”. In: *Mathematical Approaches to Biomolecular Structure and Dynamics*. Vol. 82. 1996, pp. 161–185.

[82]Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. “Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes”. In: *Journal of computational physics* 23.3 (1977), pp. 327–341.

[83]Chad W Hopkins et al. “Long-time-step molecular dynamics through hydrogen mass repartitioning”. In: *Journal of chemical theory and computation* 11.4 (2015), pp. 1864–1874.

[84]Herman JC Berendsen et al. “Molecular dynamics with coupling to an external bath”. In: *The Journal of chemical physics* 81.8 (1984), pp. 3684–3690.

[85]Hans C. Andersen. “Molecular dynamics simulations at constant pressure and/or temperature”. In: *The Journal of Chemical Physics* 72.4 (1980), pp. 2384–2393.

[86]Richard W. Pastor, Bernard R. Brooks, and Attila Szabo. “An analysis of the accuracy of Langevin and molecular dynamics algorithms”. In: *Molecular Physics* 65.6 (1988), pp. 1409–1419.

[87]Philippe H. Hünenberger. “Thermostat Algorithms for Molecular Dynamics Simulations”. In: *Advanced Computer Simulation*. Vol. 173. 2005, pp. 105–149.

[88]Keith J. Laidler. “The development of the Arrhenius equation”. In: *J. Chem. Educ.* 61.6 (1984), p. 494.

[89]Hiroaki Fukunishi, Osamu Watanabe, and Shoji Takada. “On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction”. In: *The Journal of chemical physics* 116.20 (2002), pp. 9058–9067.

[90]Nicholas Metropolis et al. “Equation of state calculations by fast computing machines”. In: *The journal of chemical physics* 21.6 (1953), pp. 1087–1092.

[91]Danial Sabri Dashti and Adrian E. Roitberg. “Optimization of Umbrella Sampling Replica Exchange Molecular Dynamics by Replica Positioning”. In: *J. Chem. Theory Comput.* 9.11 (2013), pp. 4692–4699.

[92]*Free Energy Calculations: Theory and Applications in Chemistry and Biology*. Springer Berlin Heidelberg, 2007.

[93]Gilbert N. Lewis, Merle Randall, and Kenneth S. Pitzer. *Thermodynamics: Tab.* McGraw-Hill, 1961.

[94]Ercheng Wang et al. “End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design”. In: *Chem. Rev.* 119.16 (2019), pp. 9478–9508.

[95]Tingjun Hou et al. “Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations”. In: *Journal of chemical information and modeling* 51.1 (2011), pp. 69–82.

[96]Samuel Genheden and Ulf Ryde. “The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities”. In: *Expert opinion on drug discovery* 10.5 (2015), pp. 449–461.

[97]Bill R Miller III et al. “MMPBSA. py: an efficient program for end-state free energy calculations”. In: *Journal of chemical theory and computation* 8.9 (2012), pp. 3314–3321.

[98]Huiyong Sun et al. “Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set”. In: *Phys. Chem. Chem. Phys.* 16.31 (2014), pp. 16719–16729.

[99]M. Born. “Volumen und Hydratationswärme der Ionen”. In: *Z. Physik* 1.1 (1920), pp. 45–48.

[100]Alexey Onufriev. “Chapter 7 - Implicit Solvent Models in Molecular Dynamics Simulations: A Brief Overview”. In: *Annual Reports in Computational Chemistry*. Vol. 4. 2008, pp. 125–137.

[101]Polina V. Banushkina and Sergei V. Krivov. “Optimal reaction coordinates”. In: *WIREs Computational Molecular Science* 6.6 (2016), pp. 748–763.

[102]Johannes Kästner. “Umbrella sampling”. In: *WIREs Computational Molecular Science* 1.6 (2011), pp. 932–942.

[103]Shankar Kumar et al. “The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method”. In: *Journal of computational chemistry* 13.8 (1992), pp. 1011–1021.

[104]Fangqiang Zhu and Gerhard Hummer. “Convergence and error estimation in free energy calculations using the weighted histogram analysis method”. In: *Journal of Computational Chemistry* 33.4 (2012), pp. 453–465.

[105]Maximilian Kienlein, Martin Zacharias, and Maria M Reif. “Efficient and accurate calculation of proline cis/trans isomerization free energies from Hamiltonian replica exchange molecular dynamics simulations”. In: *Structure* 31.11 (2023), pp. 1473–1484.

[106]Gunter Fischer. “Chemical aspects of peptide bond isomerisation”. In: *Chemical Society Reviews* 29.2 (2000), pp. 119–127.

[107]David E Stewart, Atom Sarkar, and John E Wampler. “Occurrence and role ofcis peptide bonds in protein structures”. In: *Journal of molecular biology* 214.1 (1990), pp. 253–260.

[108]Pierrick Craveur et al. “Cis–trans isomerization of omega dihedrals in proteins”. In: *Amino acids* 45 (2013), pp. 279–289.

[109]Daniel JA Roderer et al. “Acceleration of protein folding by four orders of magnitude through a single amino acid substitution”. In: *Scientific reports* 5.1 (2015), p. 11840.

[110]Junchao Xia and Ronald M Levy. “Molecular dynamics of the proline switch and its role in Crk signaling”. In: *The Journal of Physical Chemistry B* 118.17 (2014), pp. 4535–4545.

[111]Philipp AM Schmidpeter, Jan Rheinberger, and Crina M Nimigean. “Prolyl isomerization controls activation kinetics of a cyclic nucleotide-gated ion channel”. In: *Nature Communications* 11.1 (2020), p. 6401.

[112]Lucia Pastorino et al. “The prolyl isomerase Pin1 regulates amyloid precursor protein processing and amyloid-$\beta$ production”. In: *Nature* 440.7083 (2006), pp. 528–534.

[113]Christopher J Nelson, Helena Santos-Rosa, and Tony Kouzarides. “Proline isomerization of histone H3 regulates lysine methylation and gene expression”. In: *Cell* 126.5 (2006), pp. 905–916.

[114]Eugen Werwein et al. “Activation of the oncogenic transcription factor B-Myb via multisite phosphorylation and prolyl cis/trans isomerization”. In: *Nucleic acids research* 47.1 (2019), pp. 103–121.

[115]Stefano Stifani. “The multiple roles of peptidyl prolyl isomerases in brain cancer”. In: *Biomolecules* 8.4 (2018), p. 112.

[116]A Salahuddin. “Proline peptide isomerization and protein folding”. In: *Journal of Biosciences* 6 (1984), pp. 349–355.

[117]Michael Levitt. “Effect of proline residues on protein folding”. In: *Journal of molecular biology* 145.1 (1981), pp. 251–263.

[118]Philipp AM Schmidpeter and Franz X Schmid. “Prolyl isomerization and its catalysis in protein folding and protein function”. In: *Journal of molecular biology* 427.7 (2015), pp. 1609–1631.

[119]Barbara Eckert et al. “Prolyl isomerization as a molecular timer in phage infection”. In: *Nature structural & molecular biology* 12.7 (2005), pp. 619–623.

[120]Roman P Jakob and Franz X Schmid. “Molecular determinants of a native-state prolyl isomerization”. In: *Journal of molecular biology* 387.4 (2009), pp. 1017–1031.

[121]Yasushige Yonezawa et al. “Intra-and intermolecular interaction inducing pyramidalization on both sides of a proline dipeptide during isomerization: an ab initio QM/MM molecular dynamics simulation study in explicit water”. In: *Journal of the American Chemical Society* 131.12 (2009), pp. 4535–4540.

[122]Stefan Fischer, Roland L Dunbrack Jr, and Martin Karplus. “Cis-trans imide isomerization of the proline dipeptide”. In: *Journal of the American Chemical Society* 116.26 (1994), pp. 11931–11937.

[123]Glenn M Torrie and John P Valleau. “Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling”. In: *Journal of computational physics* 23.2 (1977), pp. 187–199.

[124]Mahmoud Moradi et al. “A classical molecular dynamics investigation of the free energy and structure of short polyproline conformers”. In: *The Journal of chemical physics* 133.12 (2010).

[125]Volodymyr Babin, Christopher Roland, and Celeste Sagui. “Adaptively biased molecular dynamics for free energy calculations”. In: *The Journal of chemical physics* 128.13 (2008).

[126]Yuji Sugita and Yuko Okamoto. “Replica-exchange molecular dynamics method for protein folding”. In: *Chemical physics letters* 314.1-2 (1999), pp. 141–151.

[127]Robert H Swendsen and Jian-Sheng Wang. “Replica Monte Carlo simulation of spin-glasses”. In: *Physical review letters* 57.21 (1986), p. 2607.

[128]Yuji Sugita and Yuko Okamoto. “Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape”. In: *Chemical Physics Letters* 329.3-4 (2000), pp. 261–270.

[129]Gavin E Crooks. “Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences”. In: *Physical Review E* 60.3 (1999), p. 2721.

[130]Alessandro Masiero et al. “The impact of proline isomerization on antigen binding and the analytical profile of a trispecific anti-HIV antibody”. In: *MAbs*. Vol. 12. 1. Taylor & Francis. 2020, p. 1698128.

[131]Donald Hamelberg, John Mongan, and J Andrew McCammon. “Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules”. In: *The Journal of chemical physics* 120.24 (2004), pp. 11919–11929.

[132]Yinglong Miao, Victoria A Feher, and J Andrew McCammon. “Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation”. In: *Journal of chemical theory and computation* 11.8 (2015), pp. 3584–3595.

[133]Juan Alcantara et al. “An unbound proline-rich signaling peptide frequently samples cis conformations in gaussian accelerated molecular dynamics simulations”. In: *Frontiers in Molecular Biosciences* 8 (2021), p. 734169.

[134]Maria Celeste Maschio et al. “Proline isomerization effects in the amyloidogenic protein $\beta$ 2-microglobulin”. In: *Physical Chemistry Chemical Physics* 23.1 (2021), pp. 356–367.

[135]Alessandro Laio and Michele Parrinello. “Escaping free-energy minima”. In: *Proceedings of the national academy of sciences* 99.20 (2002), pp. 12562–12566.

[136]Katja Ostermeir and Martin Zacharias. “Advanced replica-exchange sampling to study the flexibility and plasticity of peptides and proteins”. In: *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1834.5 (2013), pp. 847–853.

[137]Katja Ostermeir and Martin Zacharias. “Hamiltonian replica-exchange simulations with adaptive biasing of peptide backbone and side chain dihedral angles”. In: *Journal of computational chemistry* 35.2 (2014), pp. 150–158.

[138]Jacek Lubkowski et al. “The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p”. In: *Nature structural biology* 5.2 (1998), pp. 140–147.

[139]Alessandro Barducci, Massimiliano Bonomi, and Michele Parrinello. “Metadynamics”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1.5 (2011), pp. 826–843.

[140]Yuji Sugita, Akio Kitao, and Yuko Okamoto. “Multidimensional replica-exchange method for free-energy calculations”. In: *The Journal of chemical physics* 113.15 (2000), pp. 6042–6051.

[141]James A Maier et al. “ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB”. In: *Journal of chemical theory and computation* 11.8 (2015), pp. 3696–3713.

[142]David A Case et al. “The Amber biomolecular simulation programs”. In: *Journal of computational chemistry* 26.16 (2005), pp. 1668–1688.

[143]William L Jorgensen et al. “Comparison of simple potential functions for simulating liquid water”. In: *The Journal of chemical physics* 79.2 (1983), pp. 926–935.

[144]Pekka Mark and Lennart Nilsson. “Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K”. In: *The Journal of Physical Chemistry A* 105.43 (2001), pp. 9954–9960.

[145]Urmi Doshi and Donald Hamelberg. “Reoptimization of the AMBER force field parameters for peptide bond (Omega) torsions using accelerated molecular dynamics”. In: *The Journal of Physical Chemistry B* 113.52 (2009), pp. 16590–16595.

[146]Andrew R Leach. *Molecular modelling: principles and applications*. Pearson education, 2001.

[147]John Nickolls et al. “Scalable parallel programming with cuda: Is cuda the parallel programming model that application developers have been waiting for?” In: *Queue* 6.2 (2008), pp. 40–53.

[148]D.A. Case et al. “AMBER 2020, University of California, San Francisco”. In: (2020).

[149]Ramu Anandakrishnan, Boris Aguilar, and Alexey V Onufriev. “H++ 3.0: automating p K prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations”. In: *Nucleic acids research* 40.W1 (2012), W537–W541.

[150]Robert W Zwanzig. “High-temperature equation of state by a perturbation method. I. Nonpolar gases”. In: *The Journal of Chemical Physics* 22.8 (1954), pp. 1420–1426.

[152]Johannes Kraml et al. “X-Entropy: A parallelized kernel density estimator with automated bandwidth selection to calculate entropy”. In: *Journal of chemical information and modeling* 61.4 (2021), pp. 1533–1538.

[153]Daniel R Roe and Thomas E Cheatham III. “PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data”. In: *Journal of chemical theory and computation* 9.7 (2013), pp. 3084–3095.

[154]William Humphrey, Andrew Dalke, and Klaus Schulten. “VMD: visual molecular dynamics”. In: *Journal of molecular graphics* 14.1 (1996), pp. 33–38.

[155]Miklos Guttman et al. “The influence of proline isomerization on potency and stability of anti-HIV antibody 10E8”. In: *Scientific reports* 10.1 (2020), p. 14313.

[156]Henry Eyring. “The theory of absolute reaction rates”. In: *Transactions of the Faraday Society* 34 (1938), pp. 41–48.

[157]Liudmila Voronina et al. “Conformations of prolyl–peptide bonds in the bradykinin 1–5 fragment in solution and in the gas phase”. In: *Journal of the American Chemical Society* 138.29 (2016), pp. 9224–9233.

[158]Maximilian Kienlein, Martin Zacharias, and Maria M Reif. “Comprehensive Analysis of Coupled Proline Cis–Trans States in Bradykinin Using $\omega$BP-REMD Simulations”. In: *Journal of Chemical Theory and Computation* 20.6 (2024), pp. 2643–2654.

[159]B. Maigret, D. Perahia, and B. Pullman. “Molecular orbital calculations on the conformation of polypeptides and proteins IV. The conformation of the prolyl and hydroxyprolyl residues.” In: *J. Theo. Biol.* 29 (1970), pp. 275–291.

[160]Roman P Jakob and Franz X Schmid. “Energetic coupling between native-state prolyl isomerization and conformational protein folding”. In: *Journal of molecular biology* 377.5 (2008), pp. 1560–1575.

[161]H. N. Cheng and F. A. Bovey. “Cis-trans equilibrium and kinetic studies of acetyl-L-proline and glycyl-L-proline.” In: *Biopolymers* 16 (1977), pp. 1465–1472.

[162]C. Grathwohl and K. Wüthrich. “NMR studies on the rates of proline cis-trans isomerization in oligopeptides.” In: *Biopolymers* 20 (1981), pp. 2623–2633.

[163]Nicholas A Pierson et al. “Cis–trans isomerizations of proline residues are key to bradykinin conformations”. In: *Journal of the American Chemical Society* 135.8 (2013), pp. 3186–3192.

[164]Y Isaac Yang et al. “Efficient sampling over rough energy landscapes with high barriers: A combination of metadynamics with integrated tempering sampling”. In: *The Journal of Chemical Physics* 144.9 (2016).

[165]John R Cann et al. “Interaction of bradykinin with sodium dodecyl sulfate and certain acidic lipids”. In: *Peptides* 7.6 (1986), pp. 1121–1130.

[166]John R Cann et al. “A CD and an NMR study of multiple bradykinin conformations in aqueous trifluoroethanol solutions”. In: *Biopolymers: Original Research on Biomolecules* 34.7 (1994), pp. 869–878.

[167]Chiradip Chatterjee and Chaitali Mukhopadhyay. “Conformational alteration of bradykinin in presence of GM1 micelle”. In: *Biochemical and biophysical research communications* 315.4 (2004), pp. 866–871.

[168]Hideo Takeuchi and Issei Harada. “Ultraviolet resonance Raman spectroscopy of X—Proline bonds: A new marker band of hydrogen bonding at the imide C O site”. In: *Journal of Raman spectroscopy* 21.8 (1990), pp. 509–515.

[169]SUSANNIE C LEE, ANNE F RUSSELL, and WILLIAM D LAIDIG. “Three-dimensional structure of bradykinin in SDS micelles: Study using nuclear magnetic resonance, distance geometry, and restrained molecular mechanics and dynamics”. In: *International Journal of Peptide and Protein Research* 35.5 (1990), pp. 367–377.

[170]John K Young, Rickey P Hicks, and Clemens Anklin. “NMR and molecular modeling investigations of the neuropeptide substance P in the presence of 15 mM sodium dodecyl sulfate micelles”. In: *Biopolymers: Original Research on Biomolecules* 34.11 (1994), pp. 1449–1462.

[171]E. Gaggelli et al. “Calcium-binding properties and molecular organization of bradykinin. A solution ^ 1H-NMR study.” In: *Eur. J. Biochem.* 262 (1999), pp. 268–276.

[172]Gregory V Nikiforovich. “Computational molecular modeling in peptide drug design”. In: *International journal of peptide and protein research* 44.6 (1994), pp. 513–531.

[173]Joseph M Salvino, Peter R Seoane, and Roland E Dolle. “Conformational analysis of bradykinin by annealed molecular dynamics and comparison to NMR-derived conformations”. In: *Journal of computational chemistry* 14.4 (1993), pp. 438–444.

[174]Moutusi Manna and Chaitali Mukhopadhyay. “Molecular dynamics simulations of the interactions of kinin peptides with an anionic POPG bilayer”. In: *Langmuir* 27.7 (2011), pp. 3713–3722.

[175]Nicholas A Pierson et al. “Number of solution states of bradykinin from ion mobility and mass spectrometry measurements”. In: *Journal of the American Chemical Society* 133.35 (2011), pp. 13810–13813.

[176]C. F. Rodriquez et al. “Gaseous bradykinin and its singly, doubly, and triply protonated forms: A first-principles study.” In: *J. Phys. Chem.* 110 (2006), pp. 7528–7537.

[177]T. Wyttenbach, G. von Helden, and M. T. Bowers. “Gas-phase conformations of biological molecules: bradykinin”. In: *J. Am. Chem. Soc.* 118 (1996), pp. 8355–8364.

[178]Iain B Squire et al. “Bradykinin B2 receptor antagonism attenuates blood pressure response to acute angiotensin-converting enzyme inhibition in normal men”. In: *Hypertension* 36.1 (2000), pp. 132–136.

[179]Francois Marceau and Domenico Regoli. “Bradykinin receptor ligands: therapeutic perspectives”. In: *Nature reviews Drug discovery* 3.10 (2004), pp. 845–852.

[180]Devasahayam Arokiar Balaya Rex et al. “A comprehensive review on current understanding of bradykinin in COVID-19 and inflammatory diseases”. In: *Molecular Biology Reports* 49.10 (2022), pp. 9915–9927.

[181]M1 Maurer et al. “New topics in bradykinin research”. In: *Allergy* 66.11 (2011), pp. 1397–1406.

[182]G. Kotovich et al. “NMR and CD conformational studies of bradykinin and its agonists and antagonists: application to receptor binding.” In: *Biochem. Cell Biol.* 76 (1998), pp. 257–266.

[183]Jakob J Lopez et al. “The structure of the neuropeptide bradykinin bound to the human G-protein coupled receptor bradykinin B2 as determined by solid-state NMR spectroscopy”. In: *Angewandte Chemie International Edition* 47.9 (2008), pp. 1668–1671.

[184]Matthew S Glover et al. “Penultimate proline in neuropeptides”. In: *Analytical chemistry* 87.16 (2015), pp. 8466–8472.

[185]S. Piana and A. Laio. “A bias-exchange approach to protein folding.” In: *J. Phys. Chem.* 111 (2007), pp. 4553–4559.

[186]J. Pfaendtner and M. Bonomi. “Efficient sampling of high-dimensional free-energy landscapes with parallel bias metadynamics.” In: *J. Chem. Theory Comput.* 11 (2015), pp. 5062–5067.

[187]A. Gupta et al. “Exploration of high dimensional free energy landscapes by a combination of temperature-accelerated sliced sampling and parallel biasing.” In: *J. Comput. Chem.* 43 (2022), pp. 1186–1200.

[188]I. S. Joung and T.E. Cheatham III. “Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations.” In: *J. Phys. Chem. B* 112 (2008), pp. 9020–9041.

[189]Lisa Joedicke et al. “The molecular basis of subtype selectivity of human kinin G-protein-coupled receptors”. In: *Nature chemical biology* 14.3 (2018), pp. 284–290.

[190]W. C. Swope et al. “A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters.” In: *J. Chem. Phys.* 76 (1982), pp. 637–649.

[191]Fabian Pedregosa et al. “Scikit-learn: Machine learning in Python”. In: *the Journal of machine Learning research* 12 (2011), pp. 2825–2830.

[192]V. Tsui and D. A. Case. “Theory and applications of the generalized Born solvation model in macromolecular simulations.” In: *Biopolymers* 56 (2001), pp. 275–291.

[193]A. Onufriev, D. Bashford, and D. A. Case. “Exploring protein native states and large-scale conformational changes with a modified generalized Born model.” In: *Proteins* 55 (2004), pp. 383–394.

[194]J. Weiser, P. S. Shenkin, and W.C. Still. “Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO).” In: *J. Comput. Chem.* 20 (1999), pp. 217–230.

[195]H. Ottleben et al. “An NMR study of the interaction of ^15N-labelled bradykinin with an antibody mimic of the bradykinin B2 receptor.” In: *Eur. J. Biochem.* 244 (1997), p. 471.

[196]C. Bonechi et al. “Study of bradykinin conformation in the presence of model membrane by Nuclear Magnetic Resonance and molecular modelling.” In: *Biochim. Biophys. Acta* 1788 (2009), pp. 708–716.

[197]T. Richard et al. “Is the C-terminal region of bradykinin the binding site of polyphenols?” In: *J. Biomol. Struct. Dynamics* 21 (2012), pp. 379–385.

[198]M. Schubert et al. “A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ^13C chemical shift statistics.” In: *J. Biomol. NMR* 24 (2002), pp. 149–154.

[199]Maximilian Kienlein and Martin Zacharias. “Ligand binding and global adaptation of the GlnPQ substrate binding domain 2 revealed by molecular dynamics simulations”. In: *Protein Science* 29.12 (2020), pp. 2482–2494.

[200]Frederica L. Theodoulou and Ian D. Kerr. “ABC transporter research: going strong 40 years on”. In: *Biochemical Society Transactions* 43.5 (2015), pp. 1033–1040.

[201]Arthur H. F Hosie and Philip S Poole. “Bacterial ABC transporters of amino acids”. In: *Research in Microbiology* 152.3 (2001), pp. 259–270.

[202]Elie Dassa. “Natural history of ABC systems: not only transporters”. In: *Essays in Biochemistry* 50 (2011), pp. 19–42.

[203]Kaspar P Locher. “Mechanistic diversity in ATP-binding cassette (ABC) transporters”. In: *Nature structural & molecular biology* 23.6 (2016), pp. 487–493.

[204]Florence Husada et al. “Watching conformational dynamics of ABC transporters with single-molecule tools”. In: *Biochemical Society Transactions* 43.5 (Oct. 9, 2015), pp. 1041–1047.

[205]Esther Biemans-Oldehinkel, Mark K Doeven, and Bert Poolman. “ABC transporter architecture and regulatory roles of accessory domains”. In: *FEBS letters* 580.4 (2006), pp. 1023–1035.

[206]Michael L Oldham, Amy L Davidson, and Jue Chen. “Structural insights into ABC transporter mechanism”. In: *Current Opinion in Structural Biology* 18.6 (2008), pp. 726–733.

[207]Tiemen van der Heide and Bert Poolman. “ABC transporters: one, two or four extracytoplasmic substrate-binding sites?” In: *EMBO reports* (2002).

[208]Faizah Fulyani et al. “Functional diversity of tandem substrate-binding domains in ABC transporters from pathogenic bacteria”. In: *Structure* 21.10 (2013), pp. 1879–1888.

[209]Ronnie P-A Berntsson et al. “A structural classification of substrate-binding proteins”. In: *FEBS letters* 584.12 (2010), pp. 2606–2617.

[210]Giel H Scheepers, Jelger A Lycklama a Nijeholt, and Bert Poolman. “An updated structural classification of substrate-binding proteins”. In: *FEBS letters* 590.23 (2016), pp. 4393–4401.

[211]Abbas Maqbool et al. “The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity”. In: *Biochemical Society Transactions* 43.5 (2015), pp. 1011–1017.

[212]Frank C Lanfermeijer et al. “On the binding mechanism of the peptide receptor of the oligopeptide transport system of Lactococcus lactis”. In: *The EMBO journal* (2000).

[213]Gea K. Schuurman-Wolters and Bert Poolman. “Substrate Specificity and Ionic Regulation of GlnPQ from Lactococcus lactis: AN ATP-BINDING CASSETTE TRANSPORTER WITH FOUR EXTRACYTOPLASMIC SUBSTRATE-BINDING DOMAINS *”. In: *Journal of Biological Chemistry* 280.25 (2005), pp. 23785–23790.

[214]Brian H Shilton et al. “Conformational changes of three periplasmic receptors for bacterial chemotaxis and transport: the maltose-, glucose/galactose-and ribose-binding proteins.” In: *Journal of molecular biology* 264.2 (1996), pp. 350–363.

[215]Florante A. Quiocho and Polly S. Ledvina. “Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes”. In: *Molecular Microbiology* 20.1 (1996), pp. 17–25.

[216]Chun Tang, Charles D Schwieters, and G Marius Clore. “Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR”. In: *Nature* 449.7165 (2007), pp. 1078–1082.

[217]Marijn de Boer et al. “Conformational and dynamic plasticity in substrate-binding proteins underlies selective transport in ABC importers”. In: *Elife* 8 (2019), e44652.

[218]Giorgos Gouridis et al. “Conformational dynamics in substrate-binding domains influences transport in the ABC importer GlnPQ”. In: *Nature Structural & Molecular Biology* 22.1 (2015), pp. 57–64.

[219]Guillermo A. Bermejo et al. “Ligand-Free Open Closed Transitions of Periplasmic Binding Proteins: The Case of Glutamine-Binding Protein”. In: *Biochemistry* 49.9 (2010), pp. 1893–1902.

[220]Rebecca Mächtel et al. “An integrated transport mechanism of the maltose ABC importer”. In: *Research in microbiology* 170.8 (2019), pp. 321–337.

[221]Björn Hellenkamp et al. “Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study”. In: *Nat Methods* 15.9 (2018), pp. 669–676.

[222]Björn Hellenkamp et al. “Multidomain structure and correlated dynamics determined by self-consistent FRET networks”. In: *Nat Methods* 14.2 (2017), pp. 174–180.

[223]Ting Guang Sun et al. “A molecular dynamics simulation study of glutamine-binding protein”. In: *Journal of Molecular Structure: THEOCHEM* 725.1 (2005), pp. 9–16.

[224]Hannes H. Loeffler and Akio Kitao. “Collective Dynamics of Periplasmic Glutamine Binding Protein upon Domain Closure”. In: *Biophysical Journal* 97.9 (2009), pp. 2541–2549.

[225]Denis Bucher et al. “Accessing a Hidden Conformation of the Maltose Binding Protein Using Accelerated Molecular Dynamics”. In: *PLOS Computational Biology* 7.4 (2011), e1002034.

[226]Yong Wang et al. “Exploration of Multi-State Conformational Dynamics and Underlying Global Functional Landscape of Maltose Binding Protein”. In: *PLOS Computational Biology* 8.4 (2012), e1002471.

[227]Jeremy Curuksu, Jiri Sponer, and Martin Zacharias. “Elbow Flexibility of the kt38 RNA Kink-Turn Motif Investigated by Free-Energy Molecular Dynamics Simulations”. In: *Biophysical Journal* 97.7 (2009), pp. 2004–2013.

[229]Giorgos Gouridis et al. “Structural dynamics in the evolution of a bilobed protein scaffold”. In: *Proceedings of the National Academy of Sciences* 118.49 (2021), e2026165118.

[230]Maximilian Kienlein and Martin Zacharias. “How arginine inhibits substrate-binding domain 2 elucidated using molecular dynamics simulations”. In: *Protein Science* 33.7 (2024), e5077.

[231]Monika Chandravanshi, Sisir Kant Tripathi, and Shankar Prasad Kanaujia. “An updated classification and mechanistic insights into ligand binding of the substrate-binding proteins”. In: *FEBS letters* 595.18 (2021), pp. 2395–2409.

[232]Faizah Fulyani et al. “Relative rates of amino acid import via the ABC transporter GlnPQ determine the growth performance of Lactococcus lactis”. In: *Journal of bacteriology* 198.3 (2016), pp. 477–485.

[233]Justina C Wolters et al. “Ligand binding and crystal structures of the substrate-binding domain of the ABC transporter OpuA”. In: *PLoS One* 5.4 (2010), e10361.

[234]Marco Pittelkow et al. “The crystal structure of the substrate-binding protein OpuBC from Bacillus subtilis in complex with choline”. In: *Journal of molecular biology* 411.1 (2011), pp. 53–67.

[235]Kaur Manjeet et al. “Bacterial chitin binding proteins show differential substrate binding and synergy with chitinases”. In: *Microbiological research* 168.7 (2013), pp. 461–468.

[236]Chiliang Chen et al. “The ATP-binding cassette transporter Cbc (choline/betaine/-carnitine) recruits multiple substrate-binding proteins with strong specificity for distinct quaternary ammonium compounds”. In: *Molecular microbiology* 75.1 (2010), pp. 29–45.

[237]Yitao Feng et al. “Conformational dynamics of apo-GlnBP revealed by experimental and computational analysis”. In: *Angewandte Chemie International Edition* 55.45 (2016), pp. 13990–13994.

[238]Andrew J. Sharff, Lynn E. Rodseth, and Florante A. Quiocho. “Refined 1.8-.ANG. structure reveals the mode of binding of .beta.-cyclodextrin to the maltodextrin binding protein”. In: *Biochemistry* 32.40 (1993), pp. 10553–10559.

[239]Marco van den Noort, Marijn de Boer, and Bert Poolman. “Stability of ligand-induced protein conformation influences affinity in maltose-binding protein”. In: *Journal of molecular biology* 433.15 (2021), p. 167036.

[240]Martin F. Peter et al. “Triggering Closure of a Sialic Acid TRAP Transporter Substrate Binding Protein through Binding of Natural or Artificial Substrates”. In: *Journal of Molecular Biology* 433.3 (2021), p. 166756.

[241]Konstantinos Tassis et al. “Single-molecule studies of conformational states and dynamics in the ABC importer OpuA”. In: *FEBS Letters* 595.6 (2021), pp. 717–734.

[242]Evelyn Ploetz et al. “Structural and biophysical characterization of the tandem substrate-binding domains of the ABC importer GlnPQ”. In: *Open Biology* 11.4 (2021), p. 200406.

[243]Zhongying Han et al. “Dissecting mechanisms of ligand binding and conformational changes in the glutamine-binding protein”. In: *bioRxiv* (2023), pp. 2023–08.

[244]Brian H Shilton. “Active transporters as enzymes: an energetic framework applied to major facilitator superfamily and ABC importer systems”. In: *Biochemical Journal* 467.2 (2015), pp. 193–199.

[245]Jue Chen et al. “Trapping the transition state of an ATP-binding cassette transporter: evidence for a concerted mechanism of maltose transport”. In: *Proceedings of the National Academy of Sciences* 98.4 (2001), pp. 1525–1530.

[246]Cedric Orelle et al. “Both maltose-binding protein and ATP are required for nucleotide-binding domain closure in the intact maltose ABC transporter”. In: *Proceedings of the National Academy of Sciences* 105.35 (2008), pp. 12837–12842.

[247]Michael L Oldham, Shanshuang Chen, and Jue Chen. “Structural basis for substrate specificity in the Escherichia coli maltose transport system”. In: *Proceedings of the National Academy of Sciences* 110.45 (2013), pp. 18132–18137.

[248]Jie Yu et al. “Structural basis for substrate specificity of an amino acid ABC transporter”. In: *Proceedings of the National Academy of Sciences* 112.16 (2015), pp. 5243–5248.

[249]Mats HM Olsson et al. “PROPKA3: consistent treatment of internal and surface residues in empirical p K a predictions”. In: *Journal of chemical theory and computation* 7.2 (2011), pp. 525–537.

[250]Hyung-June Woo and Benoıt Roux. “Calculation of absolute protein–ligand binding free energy from computer simulations”. In: *Proceedings of the National Academy of Sciences* 102.19 (2005), pp. 6825–6830.

[251]Alan Grossfield. “An Implementation of WHAM: The Weighted Histogram Analysis Method, version 2.0. 9. 18”. In: (2014).

[252]Surjit B. Dixit et al. “Solvation thermodynamics of amino acids Assessment of the electrostatic contribution and force-field dependence”. In: *Journal of the Chemical Society, Faraday Transactions* 93.6 (1997), pp. 1105–1113.

[253]Céline Houriez, Michael Meot-Ner (Mautner), and Michel Masella. “Solvation of the Guanidinium Ion in Pure Aqueous Environments: A Theoretical Study from an “Ab Initio”-Based Polarizable Force Field”. In: *J. Phys. Chem. B* 121.50 (2017), pp. 11219–11228.

[254]Richard T. Bradshaw et al. “The Role of Electrostatics in Enzymes: Do Biomolecular Force Fields Reflect Protein Electric Fields?” In: *J. Chem. Inf. Model.* 60.6 (2020), pp. 3131–3144.

[255]Marcus Fischer et al. “Caught in a TRAP: substrate-binding proteins in secondary transport”. In: *Trends in microbiology* 18.10 (2010), pp. 471–478.

[256]Tamjeed Saleh and Charalampos G Kalodimos. “Enzymes at work are enzymes in motion”. In: *Science* 355.6322 (2017), pp. 247–248.

[257]B Poolman, E J Smid, and W N Konings. “Kinetic properties of a phosphate-bond-driven glutamate-glutamine transport system in Streptococcus lactis and Streptococcus cremoris”. In: *Journal of Bacteriology* 169.6 (1987), pp. 2755–2761.

[258]Jejoong Yoo and Aleksei Aksimentiev. “New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions”. In: *Phys. Chem. Chem. Phys.* 20.13 (2018), pp. 8432–8449.

[259]Jejoong Yoo and Aleksei Aksimentiev. “Improved Parameterization of Amine–Carboxylate and Amine–Phosphate Interactions for Molecular Dynamics Simulations Using the CHARMM and AMBER Force Fields”. In: *J. Chem. Theory Comput.* 12.1 (2016), pp. 430–443.

[260]Jiahua Deng and Qiang Cui. “Electronic Polarization Is Essential for the Stabilization and Dynamics of Buried Ion Pairs in Staphylococcal Nuclease Mutants”. In: *J. Am. Chem. Soc.* 144.10 (2022), pp. 4594–4610.

[261]Anna S. Kamenik et al. “Polarizable and non-polarizable force fields: Protein folding, unfolding, and misfolding”. In: *The Journal of Chemical Physics* 153.18 (2020), p. 185102.

[262]Drazen Petrov and Bojan Zagrovic. “Are Current Atomistic Force Fields Accurate Enough to Study Proteins in Crowded Environments?” In: *PLOS Computational Biology* 10.5 (2014), e1003638.

[263]Mark S. Miller et al. “Reparametrization of Protein Force Field Nonbonded Interactions Guided by Osmotic Coefficient Measurements from Molecular Dynamics Simulations”. In: *J. Chem. Theory Comput.* 13.4 (2017), pp. 1812–1826.

[264]Jason A. Hall et al. “Two Modes of Ligand Binding in Maltose-binding Protein ofEscherichia coli: FUNCTIONAL SIGNIFICANCE IN ACTIVE TRANSPORT *”. In: *Journal of Biological Chemistry* 272.28 (1997), pp. 17615–17622.

[265]Nikolai R. Skrynnikov et al. “Orienting domains in proteins using dipolar couplings measured by liquid-state NMR: differences in solution and crystal forms of maltodextrin binding protein loaded with $\beta$-cyclodextrin11Edited by P. E. Wright”. In: *Journal of Molecular Biology* 295.5 (2000), pp. 1265–1273.

[266]A. Rahman. “Correlations in the Motion of Atoms in Liquid Argon”. In: *Phys. Rev.* 136.2 (1964), A405–A411.

[267]Outi M. H. Salo-Ahen et al. “Molecular Dynamics Simulations in Drug Discovery and Pharmaceutical Development”. In: *Processes* 9.1 (2021), p. 71.

[268]Scott A. Hollingsworth and Ron O. Dror. “Molecular Dynamics Simulation for All”. In: *Neuron* 99.6 (2018), pp. 1129–1143.

[269]Arieh Warshel. “Molecular Dynamics Simulations of Biological Reactions”. In: *Acc. Chem. Res.* 35.6 (2002), pp. 385–395.

[270]Richard M. Venable, Andreas Krämer, and Richard W. Pastor. “Molecular Dynamics Simulations of Membrane Permeability”. In: *Chem. Rev.* 119.9 (2019), pp. 5954–5997.

[271]Soumil Y. Joshi and Sanket A. Deshmukh. “A review of advancements in coarse-grained molecular dynamics simulations”. In: *Molecular Simulation* 47.10 (2021), pp. 786–803.

[272]Wilfred F. van Gunsteren and Herman J. C. Berendsen. “Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry”. In: *Angewandte Chemie International Edition in English* 29.9 (1990), pp. 992–1023.

[273]Alan Grossfield and Daniel M. Zuckerman. “Chapter 2 Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations”. In: *Annual Reports in Computational Chemistry*. Vol. 5. 2009, pp. 23–48.

[274]Jörg Behler and Michele Parrinello. “Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces”. In: *Phys. Rev. Lett.* 98.14 (2007), p. 146401.

[275]Florian Häse, Loïc M. Roch, and Alán Aspuru-Guzik. “Next-Generation Experimentation with Self-Driving Laboratories”. In: *Trends in Chemistry* 1.3 (2019), pp. 282–291.

[276]V. Botu et al. “Machine Learning Force Fields: Construction, Validation, and Outlook”. In: *J. Phys. Chem. C* 121.1 (2017), pp. 511–522.

[277]Igor Poltavsky and Alexandre Tkatchenko. “Machine Learning Force Fields: Recent Advances and Remaining Challenges”. In: *J. Phys. Chem. Lett.* 12.28 (2021), pp. 6551–6564.

[278]Sandeep Kumar Sood and Pooja. “Quantum Computing Review: A Decade of Research”. In: *IEEE Transactions on Engineering Management* (2023), pp. 1–15.

[279]Samir N. Ajani et al. “Advancements in Computing: Emerging Trends in Computational Science with Next-Generation Computing”. In: *International Journal of Intelligent Systems and Applications in Engineering* 12.7 (2024), pp. 546–559.

[280]Yao Zhang and Qiang Ni. “Recent advances in quantum machine learning”. In: *Quantum Engineering* 2.1 (2020), e34.

[281]Francois-Xavier Theillet et al. “The alphabet of intrinsic disorder: I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins”. In: *Intrinsically Disordered Proteins* 1.1 (2013), e24360.

[282]Milan Kumar Hazra, Yishai Gilron, and Yaakov Levy. “Not Only Expansion: Proline Content and Density Also Induce Disordered Protein Conformation Compaction”. In: *Journal of Molecular Biology* 435.17 (2023), p. 168196.

[283]Alberto Vitali. “Proline-rich peptides: multifunctional bioactive molecules as new potential therapeutic drugs”. In: *Curr Protein Pept Sci* 16.2 (2015), pp. 147–162.

[284]Maria Grazia Murrali et al. “Proline Fingerprint in Intrinsically Disordered Proteins”. In: *ChemBioChem* 19.15 (2018), pp. 1625–1629.

[285]Borja Mateos et al. “The Ambivalent Role of Proline Residues in an Intrinsically Disordered Protein: From Disorder Promoters to Compaction Facilitators”. In: *Journal of Molecular Biology* 432.9 (2020), pp. 3093–3111.

[286]Magdalena Misiura and Wojciech Miltyk. “Proline-containing peptides—New insight and implications: A Review”. In: *BioFactors* 45.6 (2019), pp. 857–866.

[287]Henry Eyring. “The Activated Complex in Chemical Reactions”. In: *The Journal of Chemical Physics* 3.2 (1935), pp. 107–115.

[288]M.P. Allen and D.J. Tildesley. *Computer simulation of liquids*. Oxford University Press, New York, USA, 1987.

# Webpages

[@151]Kienlein, M. *Basic scripts to perform and analyze omegaBP-HREMD simulations with AMBER*. 2023. URL: `https://github.com/maxruedige/omegaBPHREMD`.

[@228]Alan Grossfield. *WHAM: the weighted histogram analysis method*. URL: `http://membrane.urmc.rochester.edu/wordpress/?`.

# A

# Supplementary Information for Chapter 3

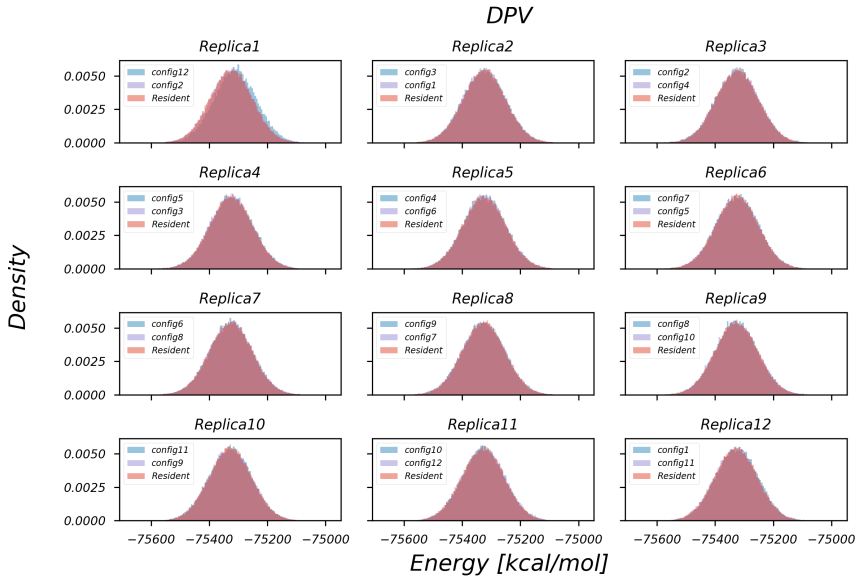



**Fig. A.1.** Probability density of potential energies of configurations in a given replica ("Resident") 1-12, compared to the probability density of potential energies evaluated for configurations in neighboring replicas ("config"$i$, where $i$ denotes the index of a neighboring replica), during a $\omega$BP-REMD simulation of 100 ns length for peptide DPV. Distribution overlap is of similar extent for the other tripeptides (data not shown).

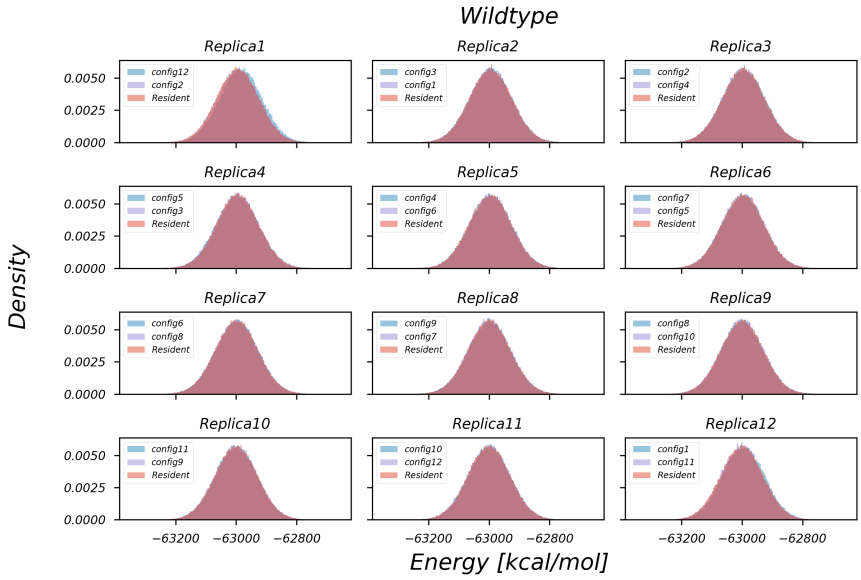



**Fig. A.2.** Probability density of potential energies of configurations in a given replica ("Resident") 1-12, compared to the probability density of potential energies evaluated for configurations in neighboring replicas ("config"$i$, where $i$ denotes the index of a neighboring replica), during a $\omega$BP-REMD simulation of 400 ns length for the wildtype N2 domain. Distribution overlap is of similar extent for the mutant versions of the N2 domain (data not shown).

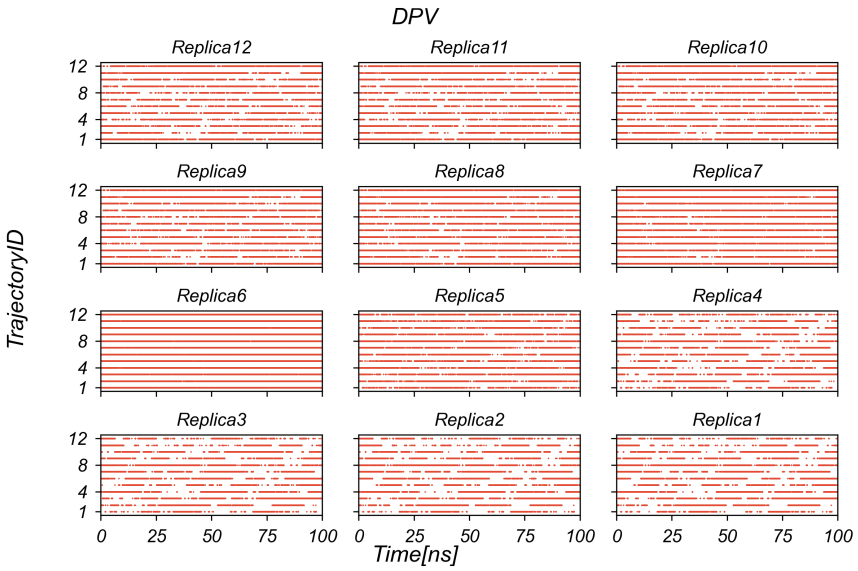



**Fig. A.3.** Sampling of trajectories 1-12 along the replica ladder during a $\omega$BP-REMD simulation of 100 ns length for peptide DPV. Configurational exchanges occur throughout all neighboring replicas. The finding also applies to the other tripeptides (data not shown).

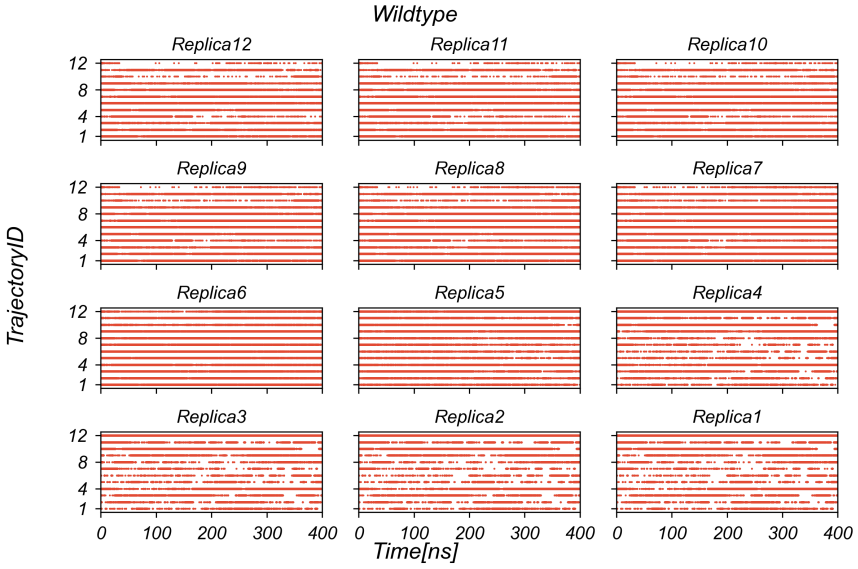



**Fig. A.4.** Sampling of trajectories 1-12 along the replica ladder during a $\omega$BP-REMD simulation of 400 ns length for the wildtype N2 domain. Configurational exchanges occur throughout all neighboring replicas. The finding also applies to the mutant versions of the N2 domain (data not shown).

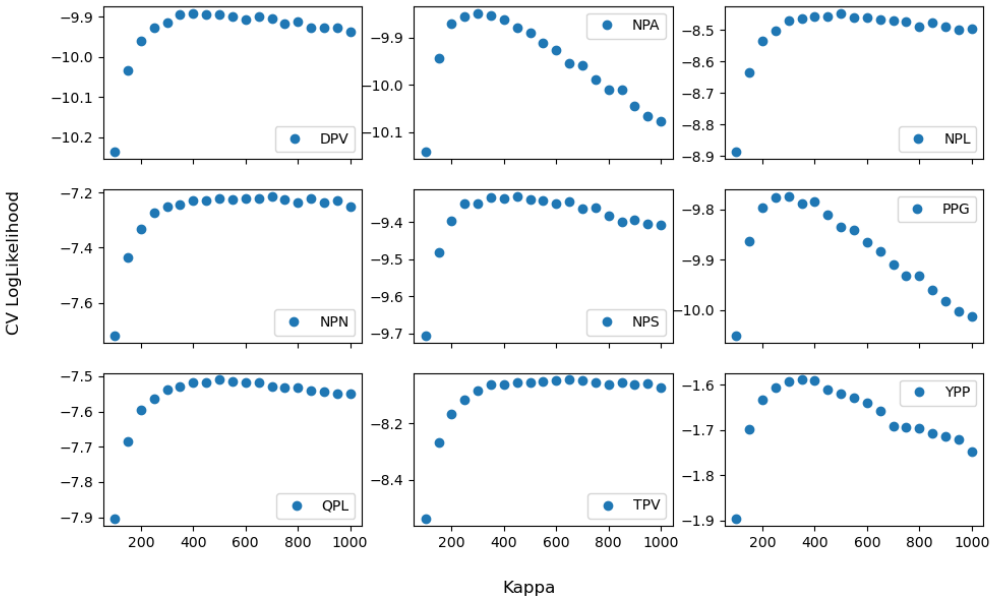



**Fig. A.5.** Maximum Likelihood Cross-Validation for all investigated tripeptide systems. The value of kappa maximizing the likelihood of the testset was used as bandwidth for the kernel density estimation.

| replica $i$ | $\alpha_{i+1}$ [%] | $\alpha_{i-1}$ [%] |
|---|---|---|
| 1 | 95.4 | - |
| 2 | 95.1 | 95.4 |
| 3 | 94.4 | 95.1 |
| 4 | 93.5 | 94.4 |
| 5 | 83.7 | 93.5 |
| 6 | 50.6 | 83.7 |
| 7 | 35.1 | 50.6 |
| 8 | 63.7 | 35.1 |
| 9 | 84.4 | 63.7 |
| 10 | 88.1 | 84.4 |
| 11 | 93.5 | 88.1 |
| 12 | - | 93.5 |

**Tab. A.1.** Replica exchange acceptance rates during a $\omega$BP-REMD simulation of 100 ns length for peptide DPV. $\alpha_{i+1}$ denotes the acceptance rate for the exchange of configurations between replica $i$ and replica $i+1$ and $\alpha_{i-1}$ denotes the acceptance rate for the exchange of configurations between replica $i$ and replica $i-1$. The acceptance rates during simulations of the other tripeptides are of similar magnitude (data not shown).

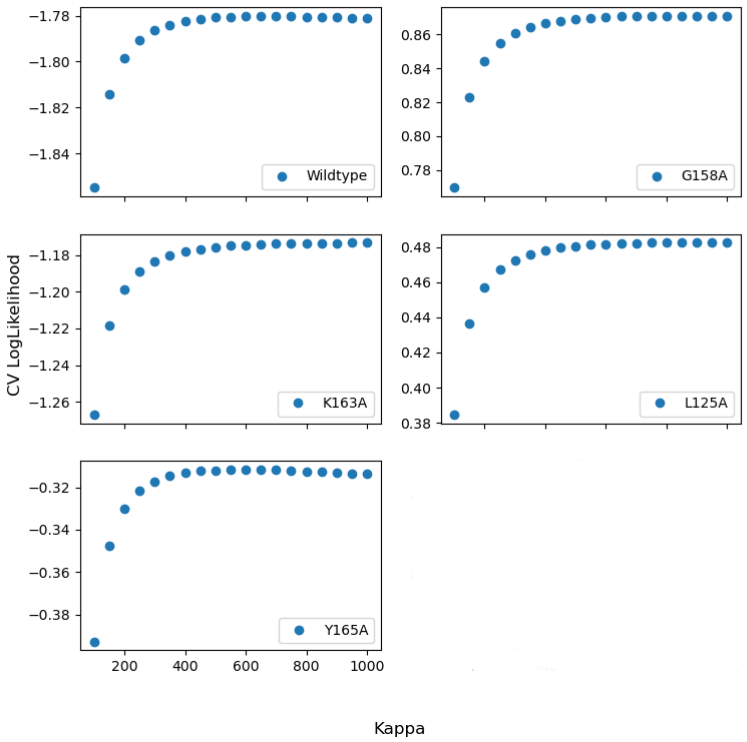



**Fig. A.6.** Maximum Likelihood Cross-Validation for all investigated variants of the N2 domain. The value of kappa maximizing the likelihood of the testset was used as bandwidth for the kernel density estimation.

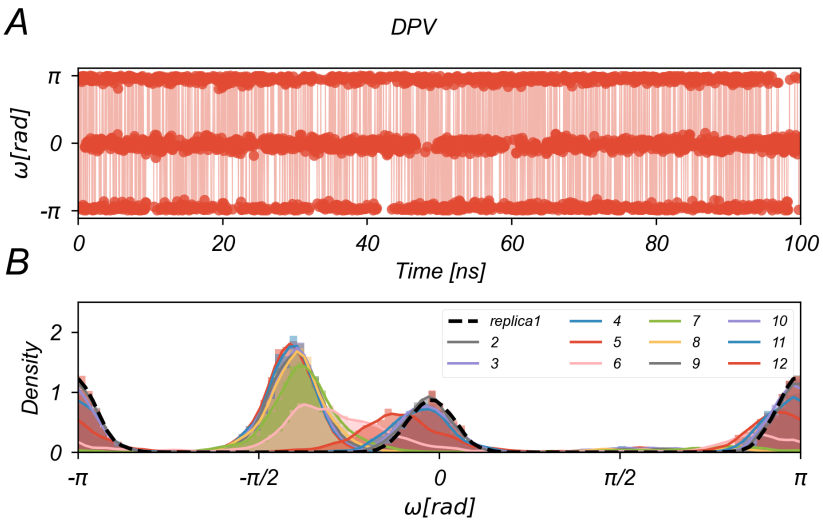



**Fig. A.7.** Cis/trans transitions observed with $\omega$BP-REMD in peptide DPV. (A) Sampling of the $\omega$ dihedral angle in replica 1 (Eq. 3.3 of the main article) during 100 ns of MD simulation. (B) Sampling in all 12 replicas depicted by the probability density along the $\omega$ dihedral angle. Exchanges via $0° \rightarrow -180°$ are preferred.

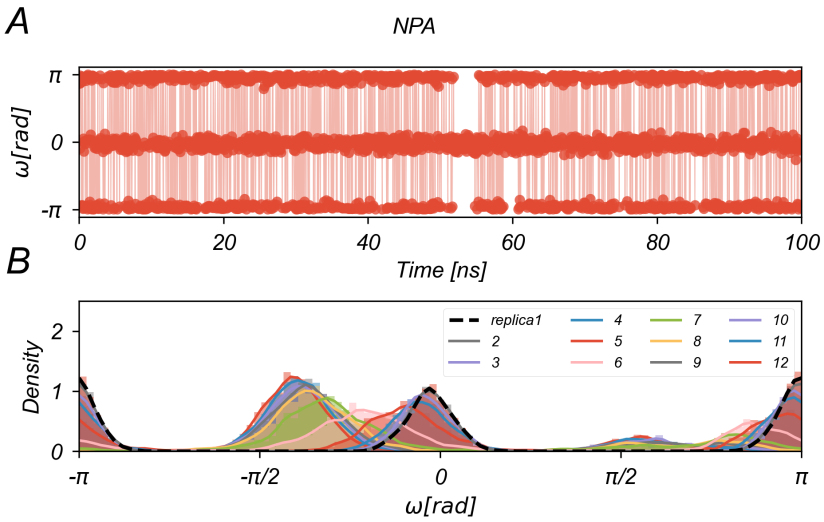



**Fig. A.8.** Cis/trans transitions observed with $\omega$BP-REMD in peptide NPA. (A) Sampling of the $\omega$ dihedral angle in replica 1 (Eq. 3.3 of the main article) during 100 ns of MD simulation. (B) Sampling in all 12 replicas depicted by the probability density along the $\omega$ dihedral angle. Exchanges via $0° \rightarrow -180°$ are preferred.

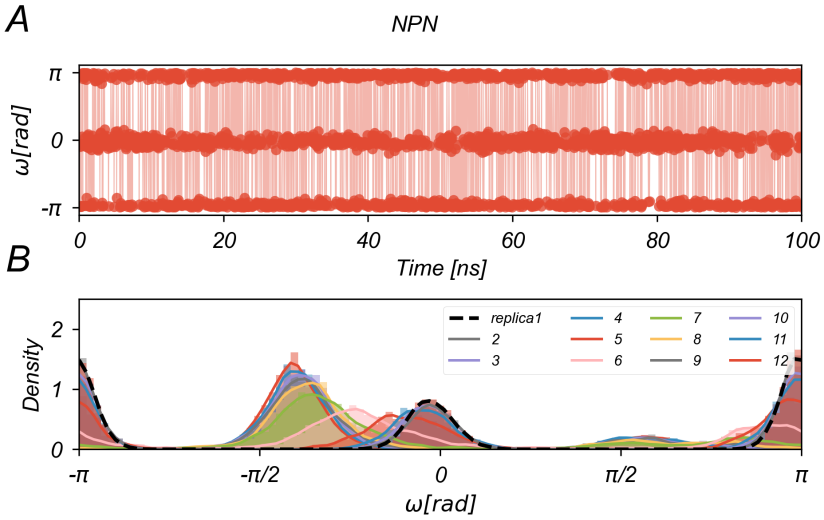



**Fig. A.9.** Cis/trans transitions observed with $\omega$BP-REMD in peptide NPN. (A) Sampling of the $\omega$ dihedral angle in replica 1 (Eq. 3.3 of the main article) during 100 ns of MD simulation. (B) Sampling in all 12 replicas depicted by the probability density along the $\omega$ dihedral angle. Exchanges via $0° \rightarrow -180°$ are preferred.

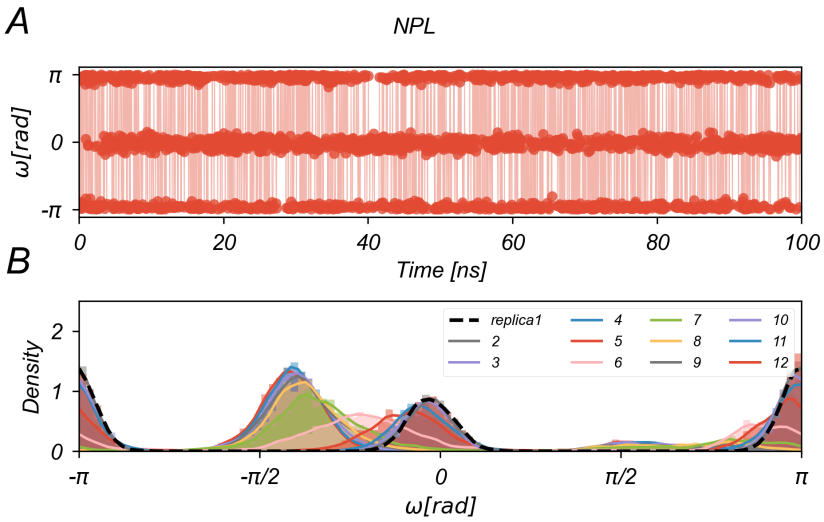



**Fig. A.10.** Cis/trans transitions observed with $\omega$BP-REMD in peptide NPL. (A) Sampling of the $\omega$ dihedral angle in replica 1 (Eq. 3.3 of the main article) during 100 ns of MD simulation. (B) Sampling in all 12 replicas depicted by the probability density along the $\omega$ dihedral angle. Exchanges via $0° \rightarrow -180°$ are preferred.

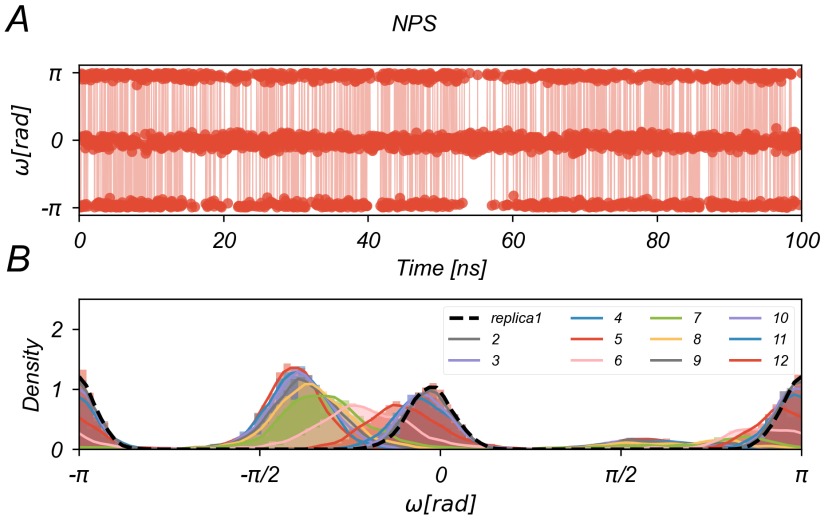



**Fig. A.11.** Cis/trans transitions observed with $\omega$BP-REMD in peptide NPS. (A) Sampling of the $\omega$ dihedral angle in replica 1 (Eq. 3.3 of the main article) during 100 ns of MD simulation. (B) Sampling in all 12 replicas depicted by the probability density along the $\omega$ dihedral angle. Exchanges via $0° \rightarrow -180°$ are preferred.

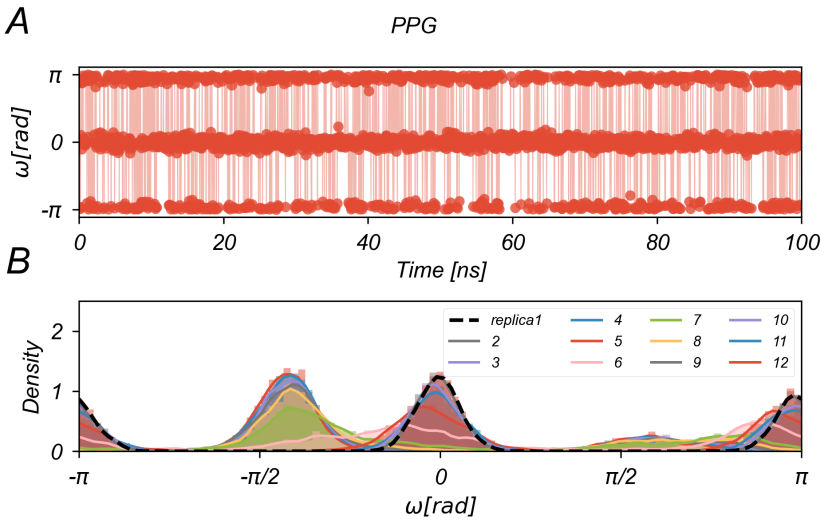



**Fig. A.12.** Cis/trans transitions observed with $\omega$BP-REMD in peptide PPG. (A) Sampling of the $\omega$ dihedral angle in replica 1 (Eq. 3.3 of the main article) during 100 ns of MD simulation. (B) Sampling in all 12 replicas depicted by the probability density along the $\omega$ dihedral angle. Exchanges via $0° \rightarrow -180°$ are preferred.

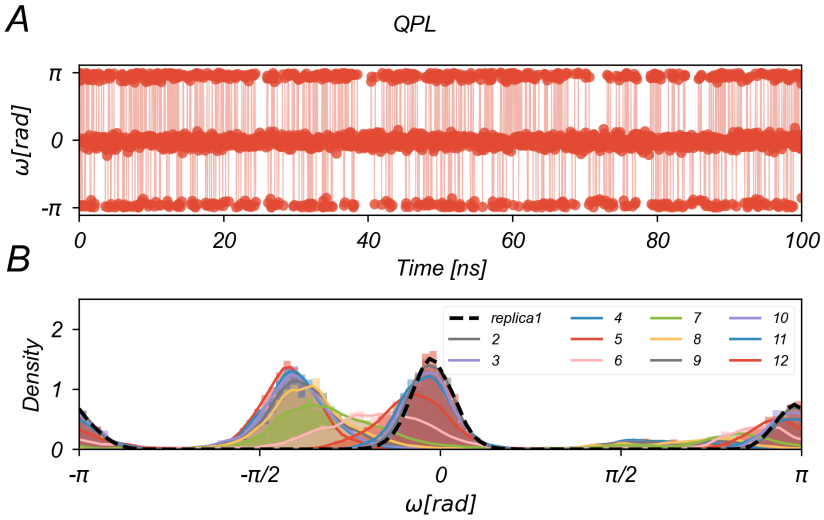



**Fig. A.13.** Cis/trans transitions observed with $\omega$BP-REMD in peptide QPL. (A) Sampling of the $\omega$ dihedral angle in replica 1 (Eq. 3.3 of the main article) during 100 ns of MD simulation. (B) Sampling in all 12 replicas depicted by the probability density along the $\omega$ dihedral angle. Exchanges via $0° \rightarrow -180°$ are preferred.

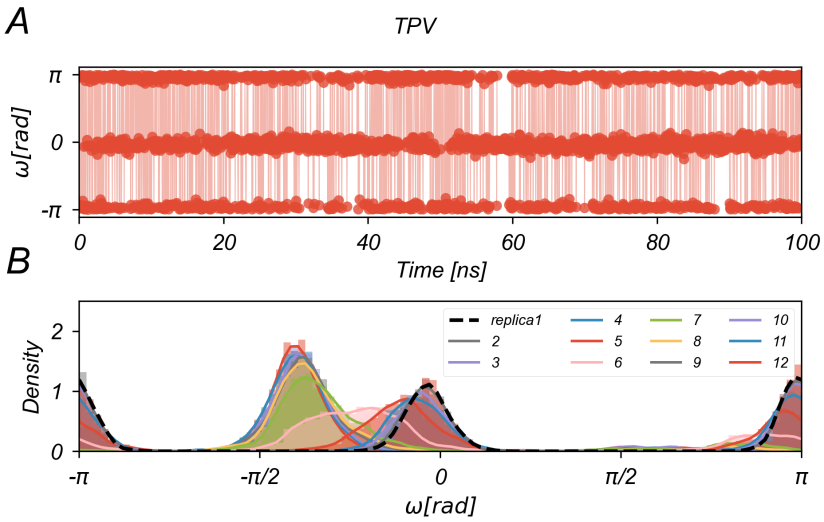



**Fig. A.14.** Cis/trans transitions observed with $\omega$BP-REMD in peptide TPV. (A) Sampling of the $\omega$ dihedral angle in replica 1 (Eq. 3.3 of the main article) during 100 ns of MD simulation. (B) Sampling in all 12 replicas depicted by the probability density along the $\omega$ dihedral angle. Exchanges via $0° \rightarrow -180°$ are preferred.

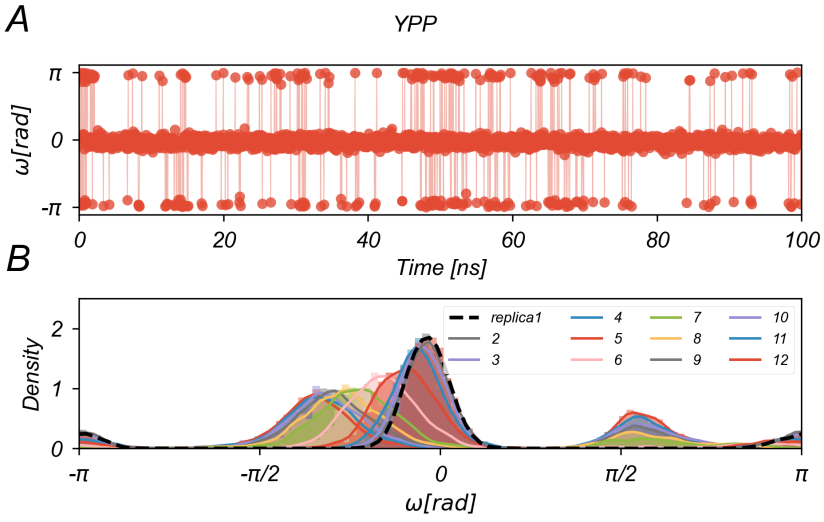



**Fig. A.15.** Cis/trans transitions observed with $\omega$BP-REMD in peptide YPP. (A) Sampling of the $\omega$ dihedral angle in replica 1 (Eq. 3.3 of the main article) during 100 ns of MD simulation. (B) Sampling in all 12 replicas depicted by the probability density along the $\omega$ dihedral angle. In contrast to the other tripeptide systems, YPP frequently samples configurations near $\pi/2$.

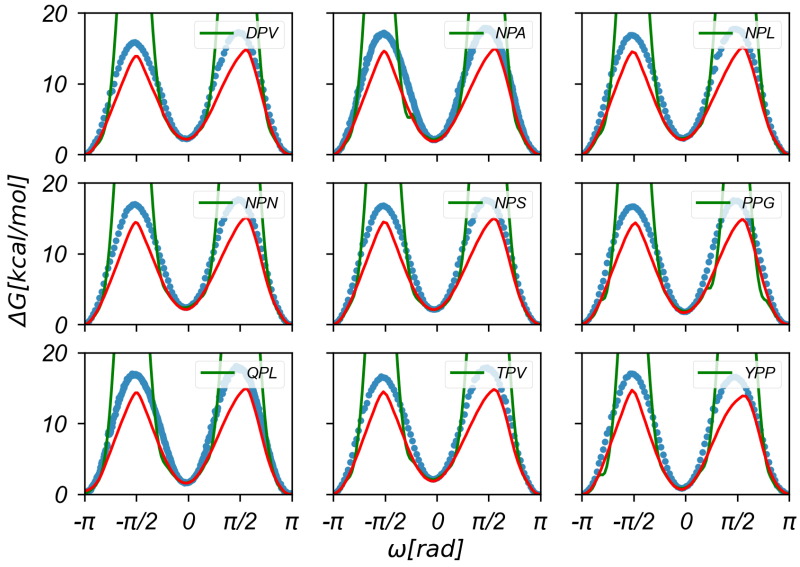



**Fig. A.16.** Different free-energy profiles along the $\omega$ dihedral angle obtained after reweighting for the tripeptide systems. Green line: Free-energy profile based on replica 1 of the $\omega$BP-REMD method. Blue dots: Free-energy profile based on all 12 replicas of the $\omega$BP-REMD method. Red line: Free-energy profile based on US. The free-energy profiles are anchored to zero in the trans state.

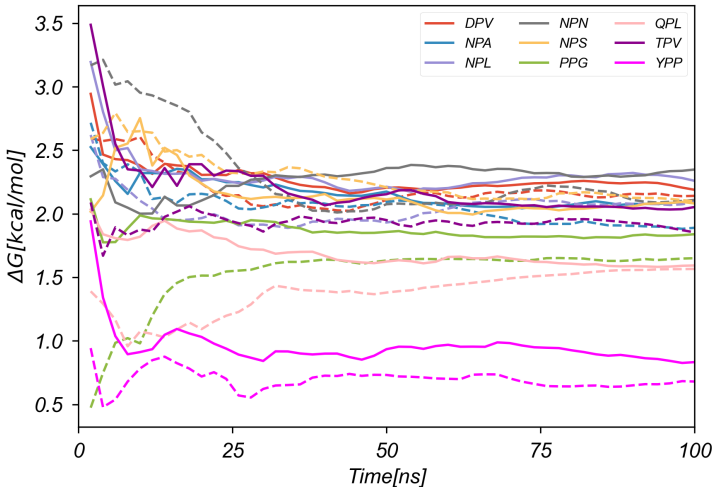



**Fig. A.17.** Convergence of the isomerization free energy $\Delta G = G_{cis} - G_{trans}$ over simulation time for all tripeptides. Solid and dashed lines correspond to the $\omega$BP-REMD and US method, respectively.

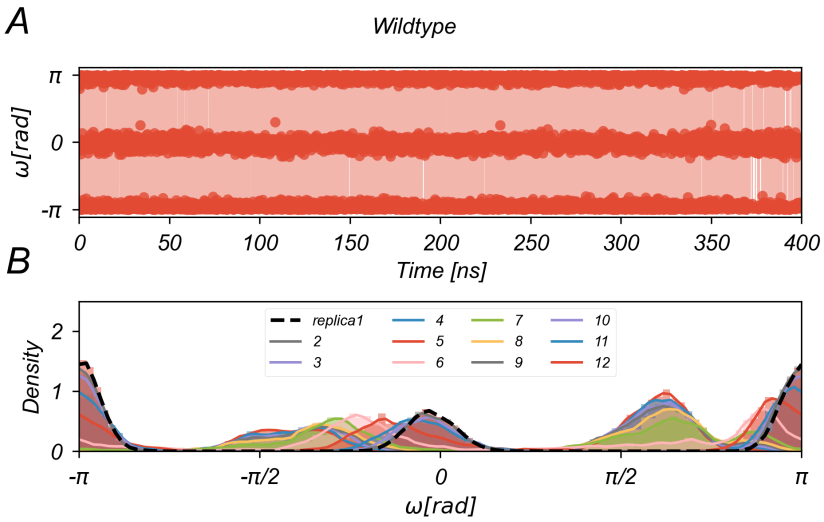



**Fig. A.18.** Cis/trans transitions observed with $\omega$BP-REMD in the wildtype N2 domain. (A) Sampling of the $\omega$ dihedral angle in replica 1 (Eq. 3.3 of the main article) during 400 ns of MD simulation. (B) Sampling in all 12 replicas depicted by the probability density along the $\omega$ dihedral angle.

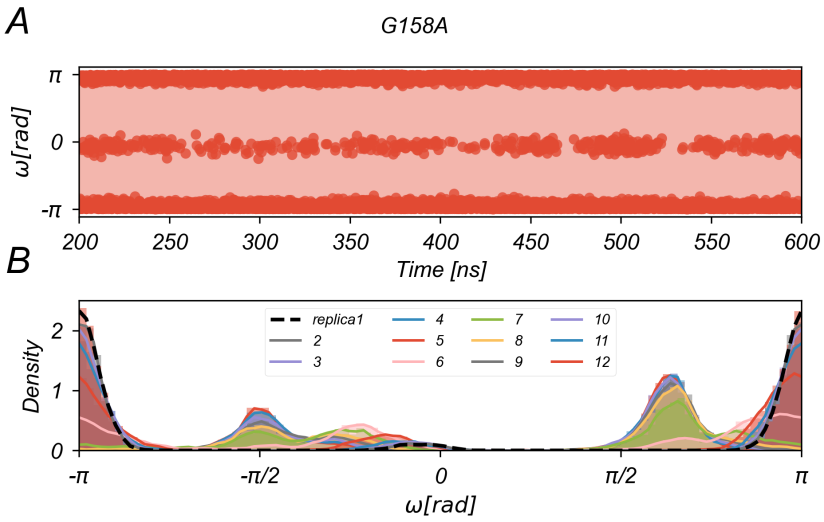



**Fig. A.19.** Cis/trans transitions observed with $\omega$BP-REMD in mutant G158A of the N2 domain. (A) Sampling of the $\omega$ dihedral angle in replica 1 (Eq. 3.3 of the main article) during 400 ns of MD simulation. The first 200 ns were used as equilibration period and are not displayed. (B) Sampling in all 12 replicas depicted by the probability density along the $\omega$ dihedral angle.

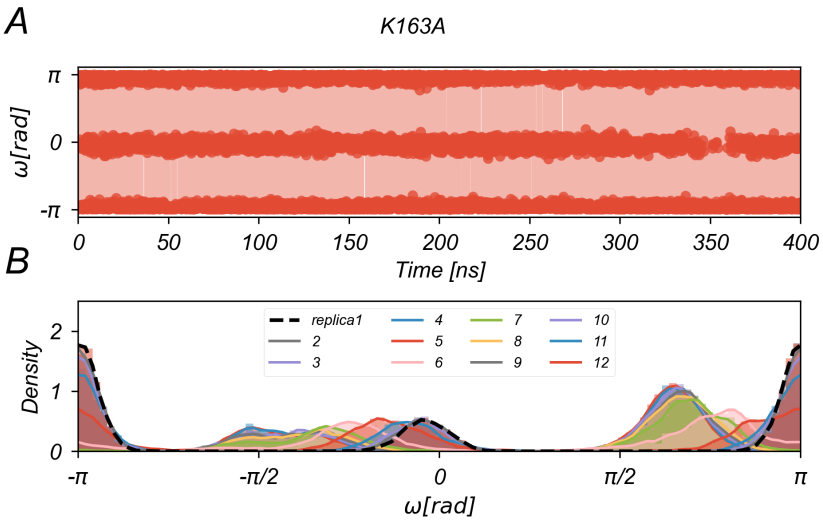



**Fig. A.20.** Cis/trans transitions observed with $\omega$BP-REMD in mutant K163A of the N2 domain. (A) Sampling of the $\omega$ dihedral angle in replica 1 (Eq. 3.3 of the main article) during 400 ns of MD simulation. (B) Sampling in all 12 replicas depicted by the probability density along the $\omega$ dihedral angle.

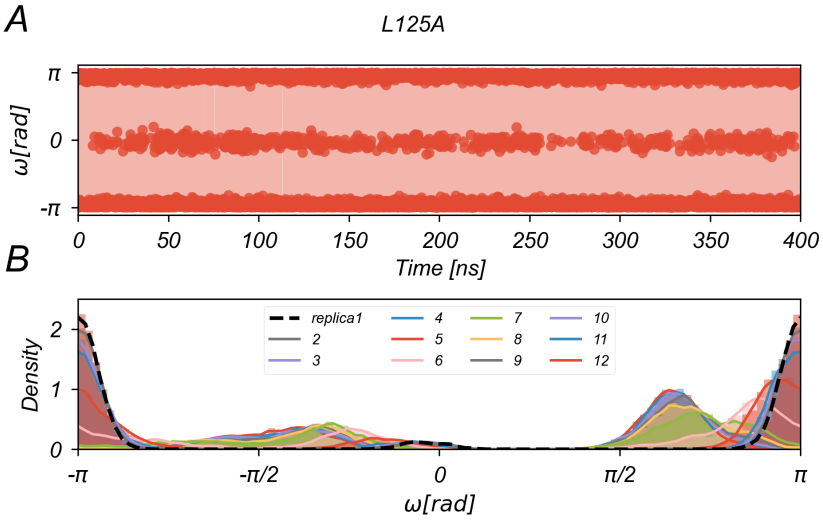



**Fig. A.21.** Cis/trans transitions observed with $\omega$BP-REMD in mutant L125A of the N2 domain. (A) Sampling of the $\omega$ dihedral angle in replica 1 (Eq. 3.3 of the main article) during 400 ns of MD simulation. (B) Sampling in all 12 replicas depicted by the probability density along the $\omega$ dihedral angle.

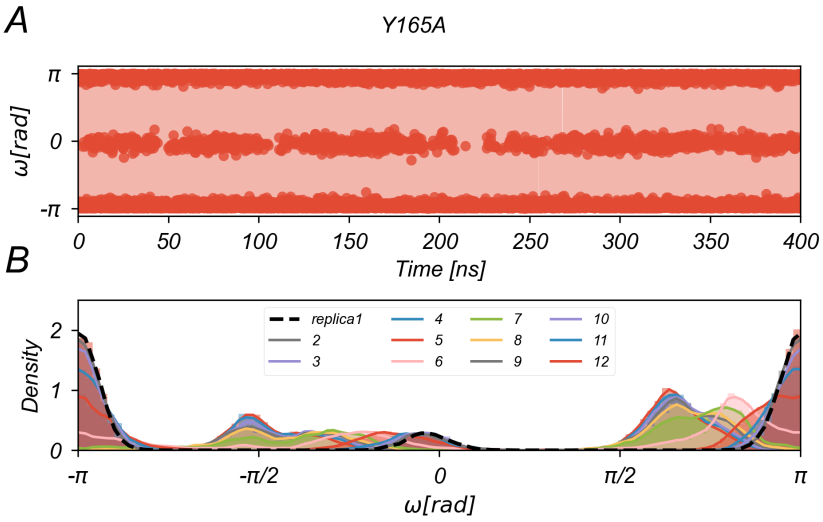



**Fig. A.22.** Cis/trans transitions observed with ωBP-REMD in mutant Y165A of the N2 domain. (A) Sampling of the $\omega$ dihedral angle in replica 1 (Eq. 3.3 of the main article) during 400 ns of MD simulation. (B) Sampling in all 12 replicas depicted by the probability density along the $\omega$ dihedral angle.

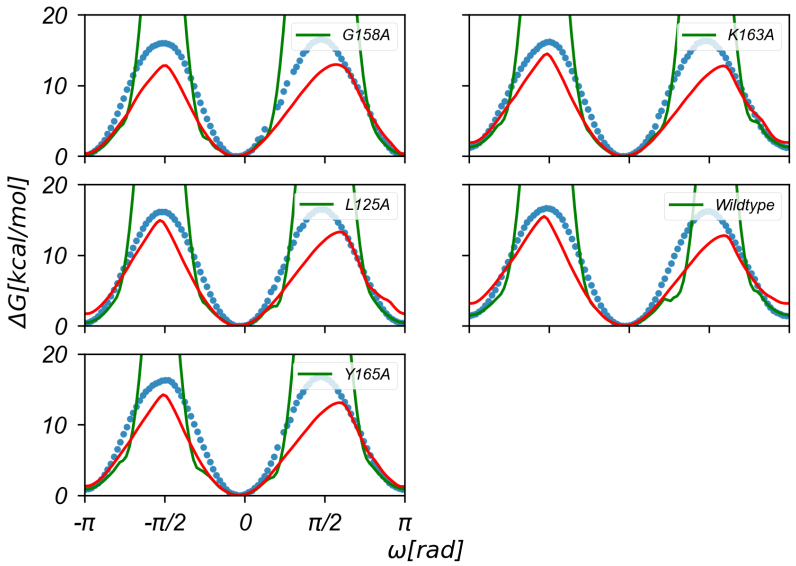



**Fig. A.23.** Different free-energy profiles along the $\omega$ dihedral angle obtained after reweighting for the protein systems. Green line: Free-energy profile based on replica 1 of the ωBP-REMD method. Blue dots: Free-energy profile based on all 12 replicas of the ωBP-REMD method. Red line: Free-energy profile based on US. The free-energy profiles are anchored to zero in the cis state.

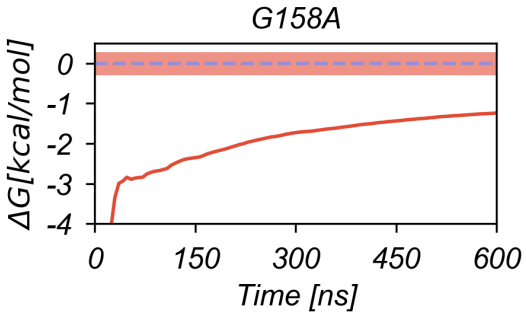



**Fig. A.24.** Isomerization free energy $\Delta G = G_{cis} - G_{trans}$ calculated based on the full 600 ns simulation time for mutant G158A of the N2 domain. Experimental values are depicted with the blue dashed line and the range of $\pm k_B T$ is shaded in transparent red. The first 200 ns were used for equilibration. Longer equilibration was necessary due to difficulties of correct placement of the additional alanine methyl group.

| replica $i$ | $\alpha_{i+1}$ [%] | $\alpha_{i-1}$ [%] |
|---|---|---|
| 1 | 95.2 | - |
| 2 | 94.8 | 95.2 |
| 3 | 94.3 | 94.8 |
| 4 | 93.6 | 94.3 |
| 5 | 84.5 | 93.6 |
| 6 | 59.6 | 84.5 |
| 7 | 38.8 | 59.6 |
| 8 | 49.9 | 38.8 |
| 9 | 77.6 | 49.9 |
| 10 | 84.6 | 77.6 |
| 11 | 91.9 | 84.6 |
| 12 | - | 91.9 |

**Tab. A.2.** Replica exchange acceptance rates during a $\omega$BP-REMD simulation of 400 ns length for the wildtype N2 domain. $\alpha_{i+1}$ denotes the acceptance rate for the exchange of configurations between replica $i$ and replica $i+1$ and $\alpha_{i-1}$ denotes the acceptance rate for the exchange of configurations between replica $i$ and replica $i-1$. The acceptance rates during simulations of the mutant versions of the N2 domain are of similar magnitude (data not shown).

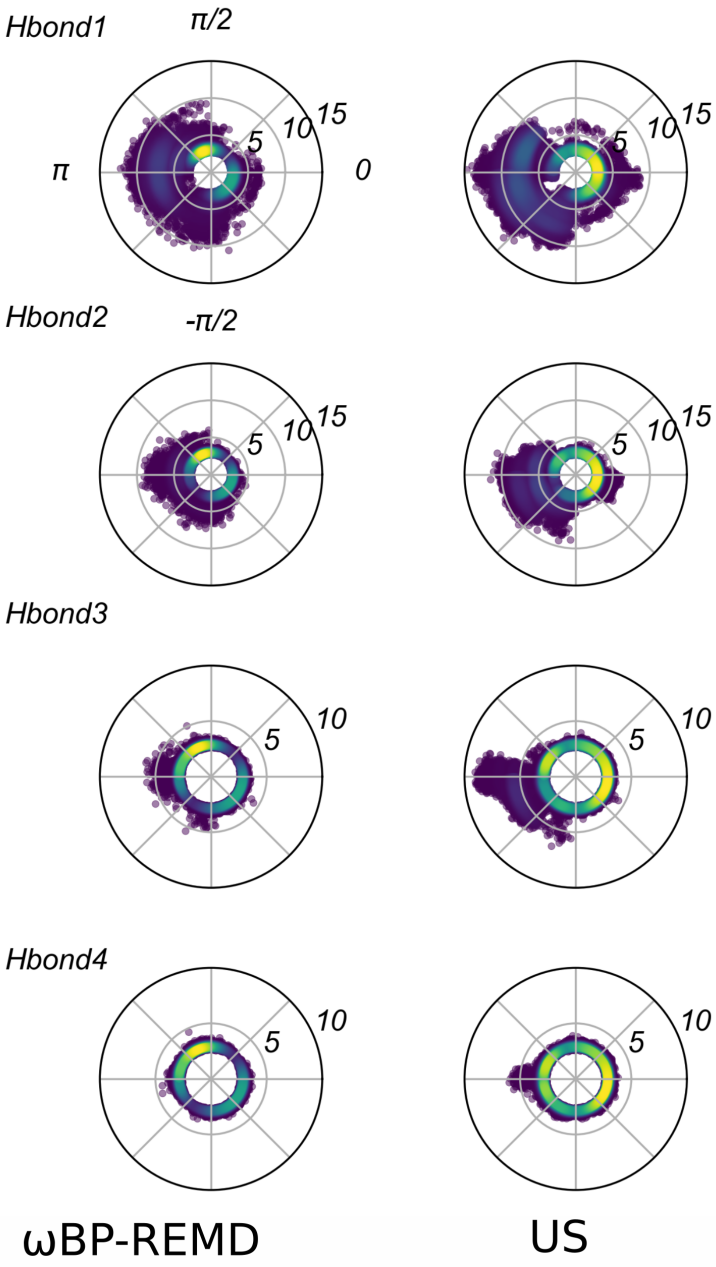



**Fig. A.25.** Analysis of hydrogen bond stability in the wildtype N2 domain. Donor-acceptor heavy atom distances in Å are plotted radially over the $\omega$ dihedral angle for hydrogen bonds 1-4 (see Fig. 3.8 of the main article for the labeling of hydrogen bonds).

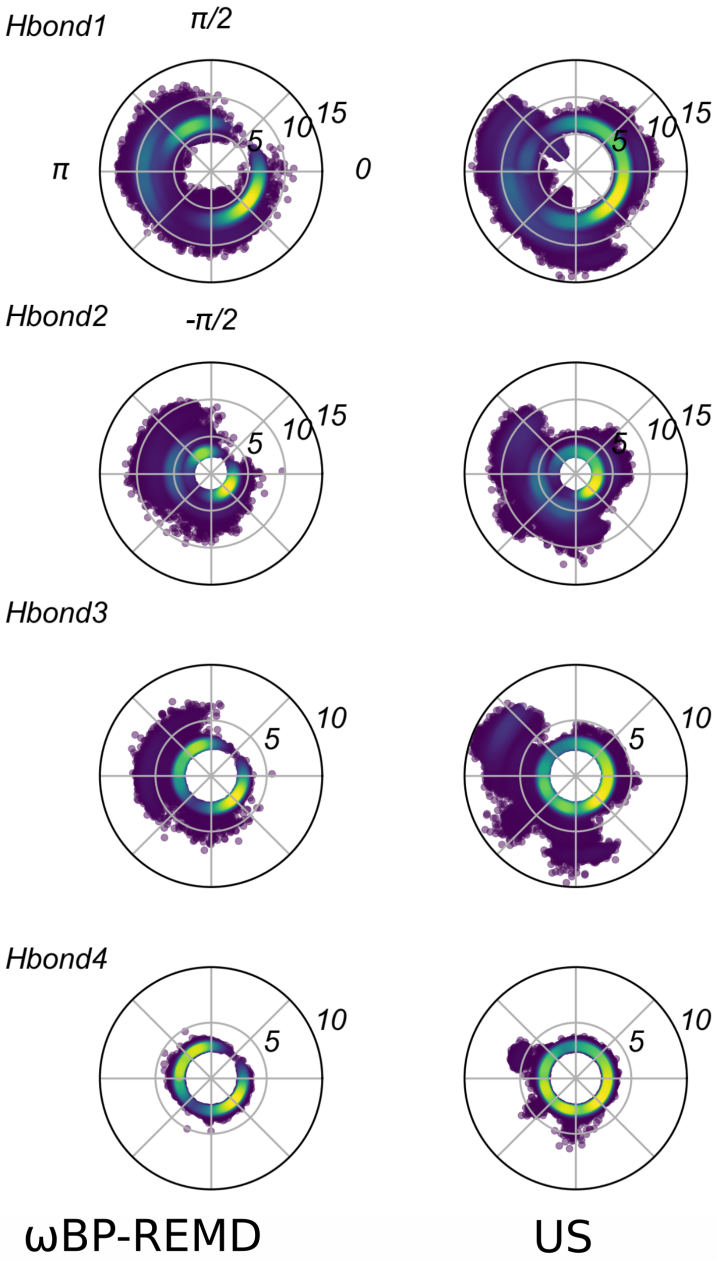



**Fig. A.26.** Analysis of hydrogen bond stability in mutant G158A of the N2 domain. Donor-acceptor heavy atom distances in Å are plotted radially over the $\omega$ dihedral angle for hydrogen bonds 1-4 (see Fig. 3.8 of the main article for the labeling of hydrogen bonds).

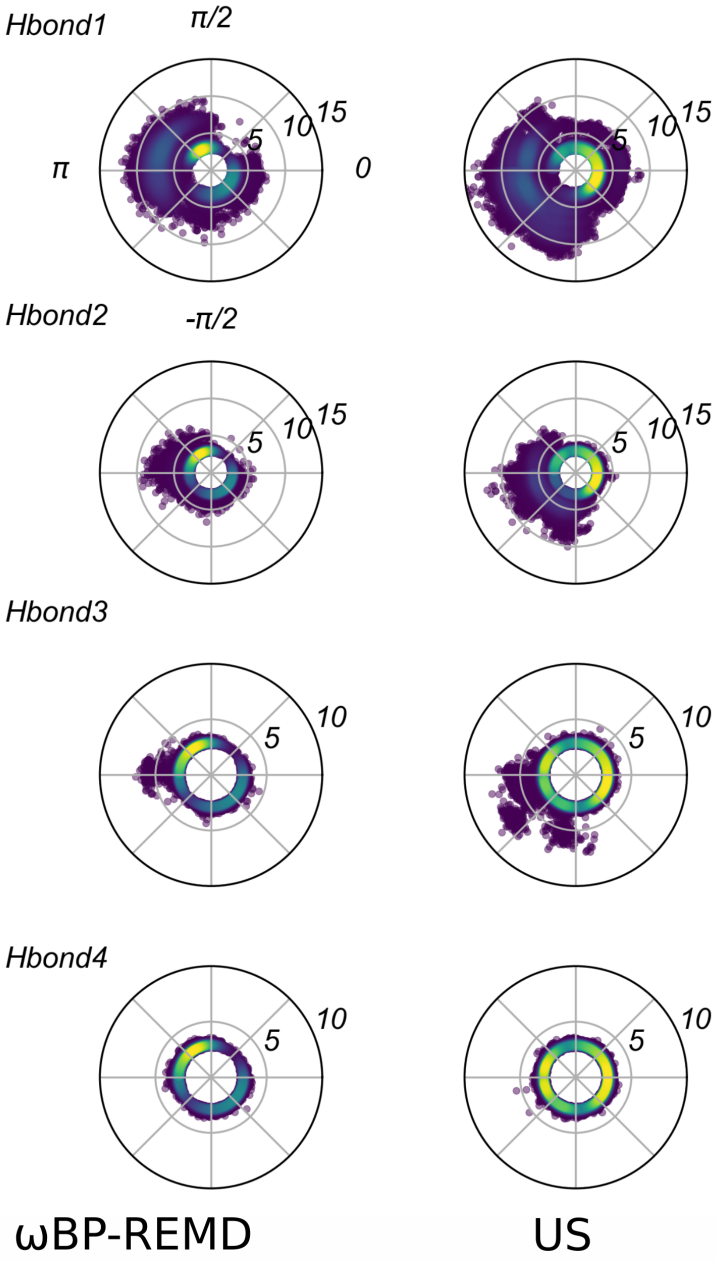



**Fig. A.27.** Analysis of hydrogen bond stability in mutant K163A of the N2 domain. Donor-acceptor heavy atom distances in Å are plotted radially over the $\omega$ dihedral angle for hydrogen bonds 1-4 (see Fig. 3.8 of the main article for the labeling of hydrogen bonds).

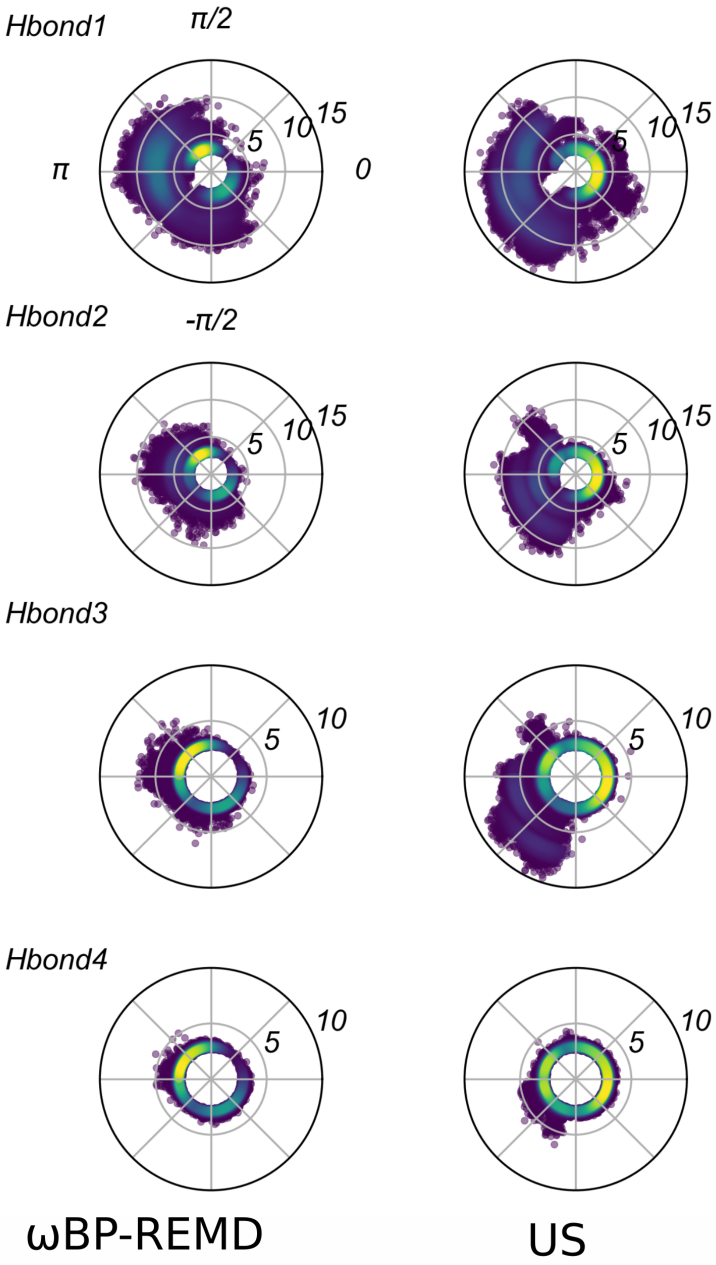



**Fig. A.28.** Analysis of hydrogen bond stability in mutant L125A of the N2 domain. Donor-acceptor heavy atom distances in Å are plotted radially over the $\omega$ dihedral angle for hydrogen bonds 1-4 (see Fig. 3.8 of the main article for the labeling of hydrogen bonds).

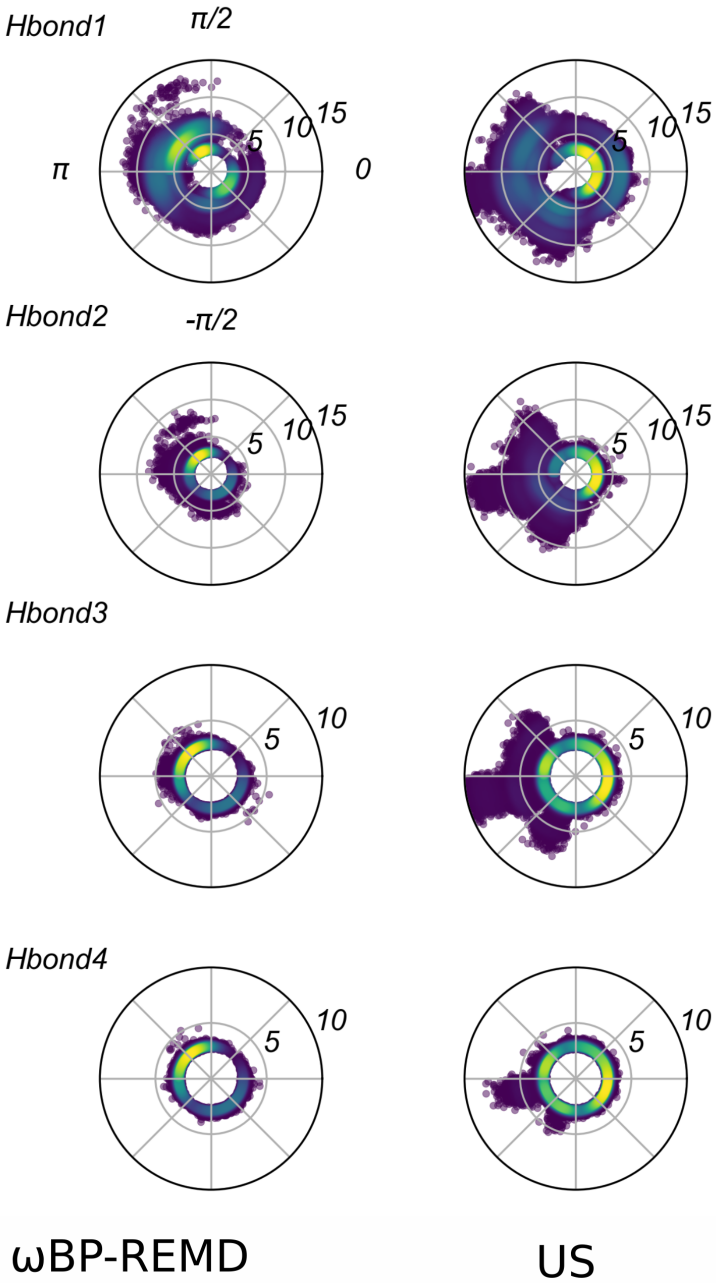



**Fig. A.29.** Analysis of hydrogen bond stability in mutant Y165A of the N2 domain. Donor-acceptor heavy atom distances in Å are plotted radially over the $\omega$ dihedral angle for hydrogen bonds 1-4 (see Fig. 3.8 of the main article for the labeling of hydrogen bonds).

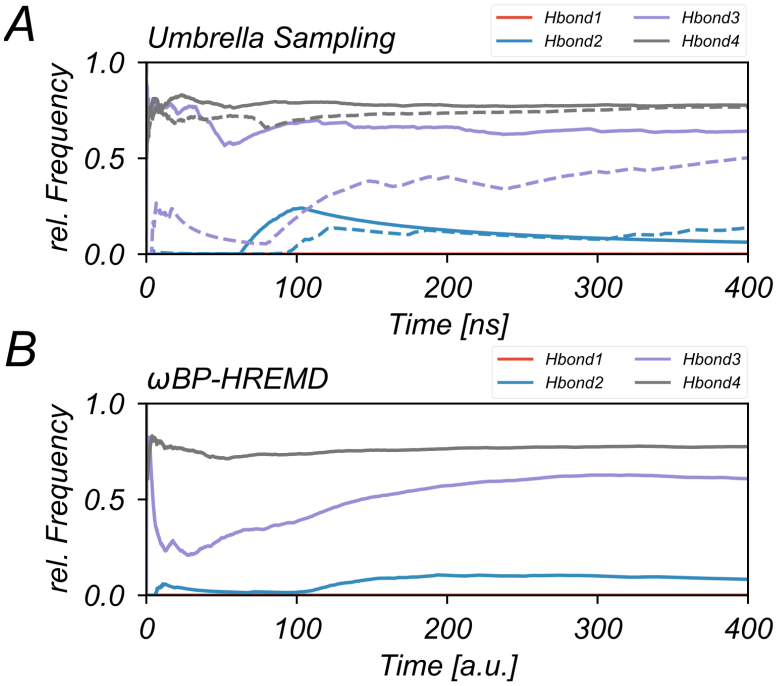



**Fig. A.30.** Relative frequency of hydrogen bond occurrence in the trans state along simulation time for mutant G158A of the N2 domain. In the upper panel the data for US is shown. The dashed lines are based on the umbrella window anchored at $\pi$ and the continuous lines based on the one anchored at $-\pi$. For the $\omega$BP-REMD method, time is given in arbitrary units, since the trans configurations were filtered out from the trajectory of replica 1, which samples both cis and trans states.

| | $\Delta G^{\ddagger}_{cis\to trans}$ [kcal·mol$^{-1}$] | $k_{cis\to trans}$ [s$^{-1}$] | $\Delta G^{\ddagger}_{trans\to cis}$ [kcal·mol$^{-1}$] | $k_{trans\to cis}$ [s$^{-1}$] |
|---|---|---|---|---|
| $\omega$BP-REMD | 16.3$^{b}$, 16.7$^{c}$ | 2.6$^{b}$, 1.3$^{c}$ | 15.1$^{c}$, 14.7$^{b}$ | 20.9$^{c}$, 42.0$^{b}$ |
| US | 12.8$^{b}$, 15.5$^{c}$ | 1161.0$^{b}$, 10.4$^{c}$ | 12.4$^{c}$, 9.5$^{b}$ | 2335.5$^{c}$, $3.7 \cdot 10^{5}$ $^{b}$ |
| exp.$^{a}$ | 12.96 | 880 | 14.2 | 100 |

**Tab. A.3.** Transition barrier height free energies $\Delta G^{\ddagger}_{cis\to trans}$ and $\Delta G^{\ddagger}_{trans\to cis}$ and corresponding rates $k_{cis\to trans}$ and $k_{trans\to cis}$ for the wildtype N2 domain, from $\omega$BP-REMD, US and experiment. The two quantities are assumed to be related based on transition state theory [287] for a unimolecular reaction $i$, $k_i = \kappa(\beta h)^{-1}\exp[-\beta\Delta G^{\ddagger}_i]$, where $\kappa$ is the transmission coefficient, $h$ is Planck's constant, $\beta = (k_B T)^{-1}$, $k_B$ is Boltzmann's constant and $T$ the absolute temperature. Here, $T = 288$ K. The transmission coefficient was assumed to be equal to unity as a first approximation because no further information on transition kinetics is available in the present work.
(a) Experimental rates are taken from Ref. [120] and the corresponding free energies are calculated from the above relation. (b) Quantity refers to the transition via $\omega = +\pi/2$ in the corresponding PMF (Fig. 3.7 of the main article). (c) Quantity refers to the transition via $\omega = -\pi/2$ in the corresponding PMF (Fig. 3.7 of the main article).

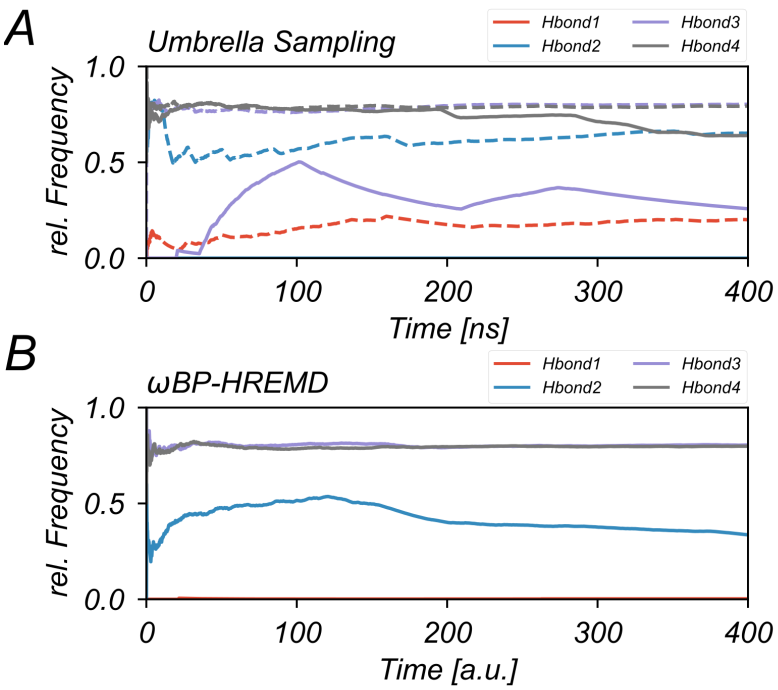



**Fig. A.31.** Relative frequency of hydrogen bond occurrence in the trans state along simulation time for the wildtype N2 domain. In the upper panel the data for US is shown. The dashed lines are based on the umbrella window anchored at $\pi$ and the continuous lines based on the one anchored at $-\pi$. For the $\omega$BP-REMD method, time is given in arbitrary units, since the trans configurations were filtered out from the trajectory of replica 1, which samples both cis and trans states.

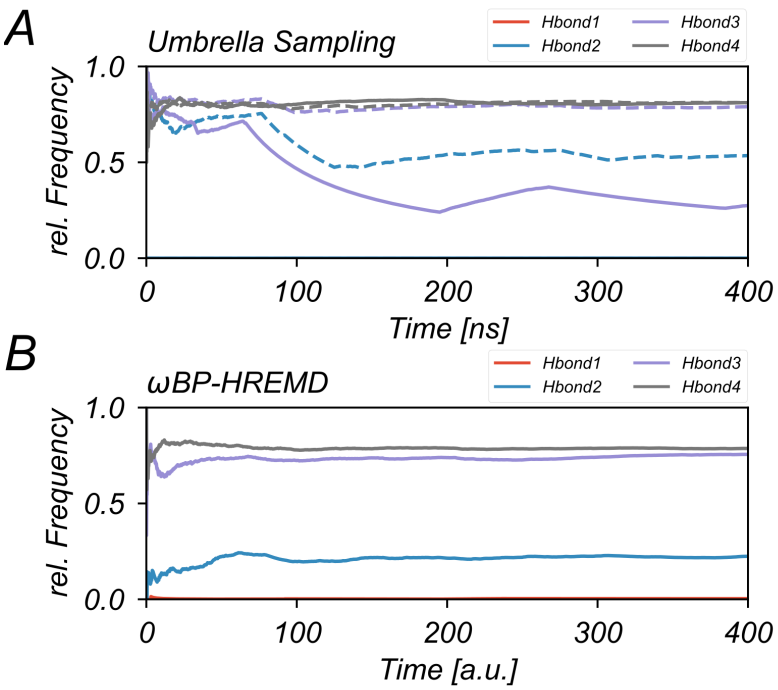



**Fig. A.32.** Relative frequency of hydrogen bond occurrence in the trans state along simulation time for mutant L125A of the N2 domain. In the upper panel the data for US is shown. The dashed lines are based on the umbrella window anchored at $\pi$ and the continuous lines based on the one anchored at $-\pi$. For the $\omega$BP-REMD method, time is given in arbitrary units, since the trans configurations were filtered out from the trajectory of replica 1, which samples both cis and trans states.

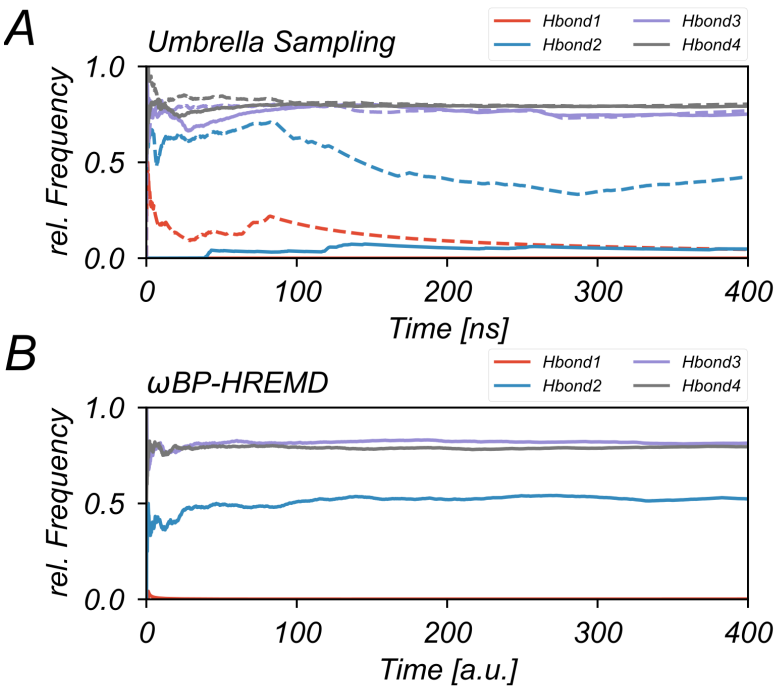



**Fig. A.33.** Relative frequency of hydrogen bond occurrence in the trans state along simulation time for mutant K163A of the N2 domain. In the upper panel the data for US is shown. The dashed lines are based on the umbrella window anchored at $\pi$ and the continuous lines based on the one anchored at $-\pi$. For the $\omega$BP-REMD method, time is given in arbitrary units, since the trans configurations were filtered out from the trajectory of replica 1, which samples both cis and trans states.

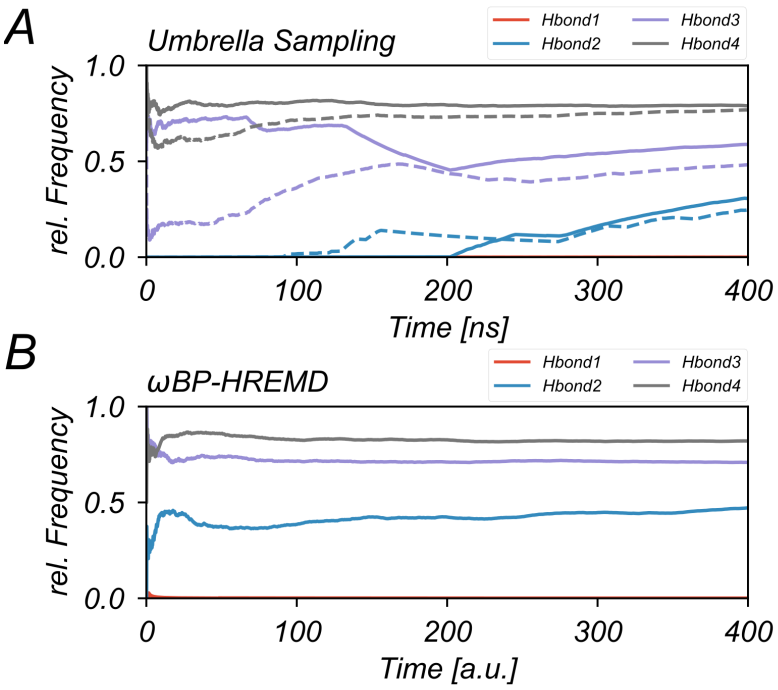



**Fig. A.34.** Relative frequency of hydrogen bond occurrence in the trans state along simulation time for mutant Y165A of the N2 domain. In the upper panel the data for US is shown. The dashed lines are based on the umbrella window anchored at $\pi$ and the continuous lines based on the one anchored at $-\pi$. For the $\omega$BP-REMD method, time is given in arbitrary units, since the trans configurations were filtered out from the trajectory of replica 1, which samples both cis and trans states.

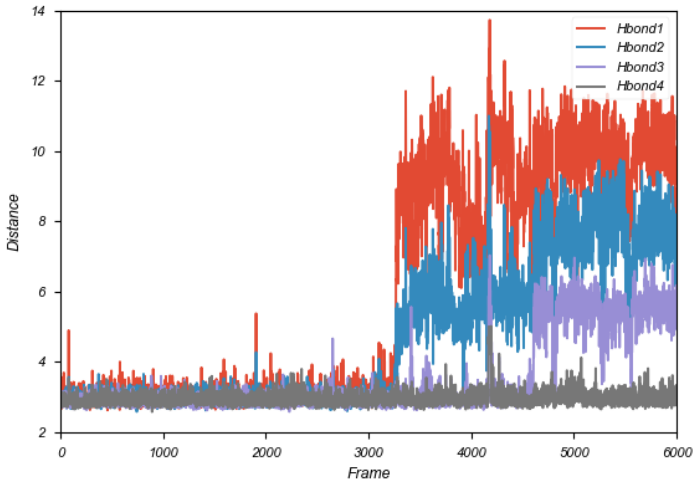



**Fig. A.35.** Donor-acceptor heavy atom distances in Å for hydrogen bonds 1-4 (see Fig. 3.8 of the main article for the labeling of hydrogen bonds) in the wildtype N2 domain for the umbrella window anchored at $\pi$.

# B

# Supplementary Information for Chapter 4

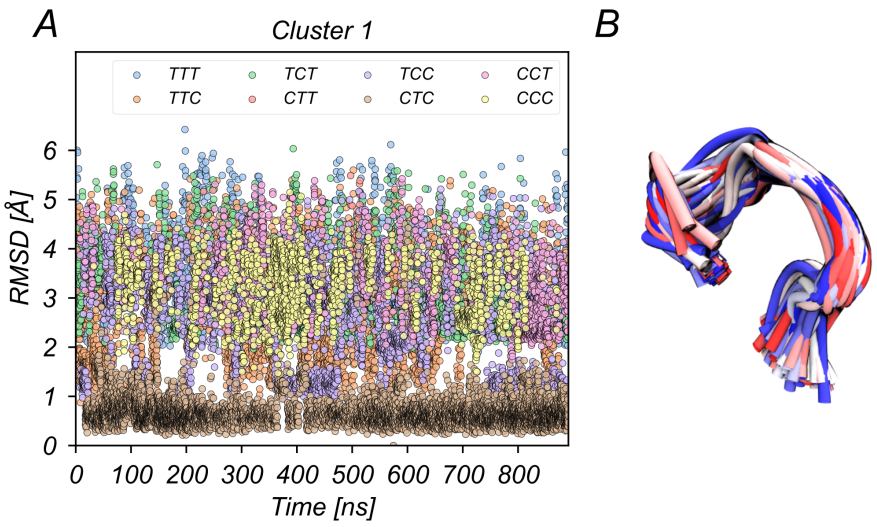



**Fig. B.1.** Characterization of the eight different isomerization states of bradykinin (BK) with respect to sampling in cluster 1 and illustration of structures sampled in cluster 1. The underlying clustering was done in a combined fashion on the entire trajectory of replica 1. (A) Root mean square deviation of the BK $C_\alpha$ atoms with respect to the configuration closest to the centroid of cluster 1, evaluated after a rototranslational fit to these atoms. The data is depicted separately for the eight BK isomers TTT, TTC, TCT, CTT, TCC, CTC, CCT and CCC. Here, the first, second and third letters refer to the trans ("T") or cis ("C") states of Pro2, Pro3, and Pro7, respectively. The trajectories pertaining to these isomerization states were filtered out from replica 1 of the $\omega$BP-REMD simulation and contained 3236, 3687, 1845, 92, 3565, 6857, 1353 and 1642 frames, respectively. Each dot represents a simulation frame. (B) Backbone superposition of the members of cluster 1. For clarity of visual illustration, only every $20^{th}$ frame along the trajectory containing the members of this cluster is shown. Colorcoding from blue to red indicates the timestep of the trajectory (dark blue, white and dark red indicating the beginning, middle and end of the trajectory, respectively).

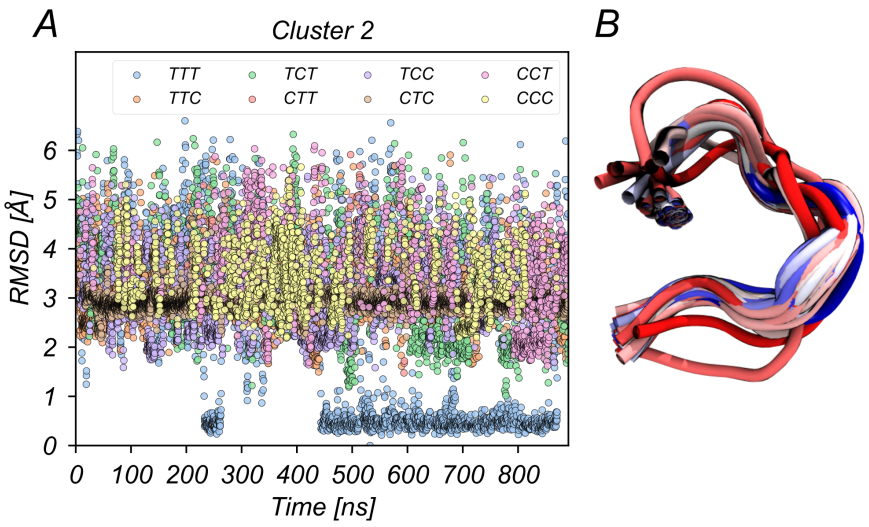



**Fig. B.2.** Characterization of the eight different isomerization states of bradykinin (BK) with respect to sampling in cluster 2 and illustration of structures sampled in cluster 2. The underlying clustering was done in a combined fashion on the entire trajectory of replica 1. (A) Root mean square deviation of the BK $C_\alpha$ atoms with respect to the configuration closest to the centroid of cluster 2, evaluated after a rototranslational fit to these atoms. The data is depicted separately for the eight BK isomers TTT, TTC, TCT, CTT, TCC, CTC, CCT and CCC. Here, the first, second and third letters refer to the trans ("T") or cis ("C") states of Pro2, Pro3, and Pro7, respectively. The trajectories pertaining to these isomerization states were filtered out from replica 1 of the $\omega$BP-REMD simulation and contained 3236, 3687, 1845, 92, 3565, 6857, 1353 and 1642 frames, respectively. Each dot represents a simulation frame. (B) Backbone superposition of the members of cluster 2. For clarity of visual illustration, only every $10^{th}$ frame along the trajectory containing the members of this cluster is shown. Colorcoding from blue to red indicates the timestep of the trajectory (dark blue, white and dark red indicating the beginning, middle and end of the trajectory, respectively).

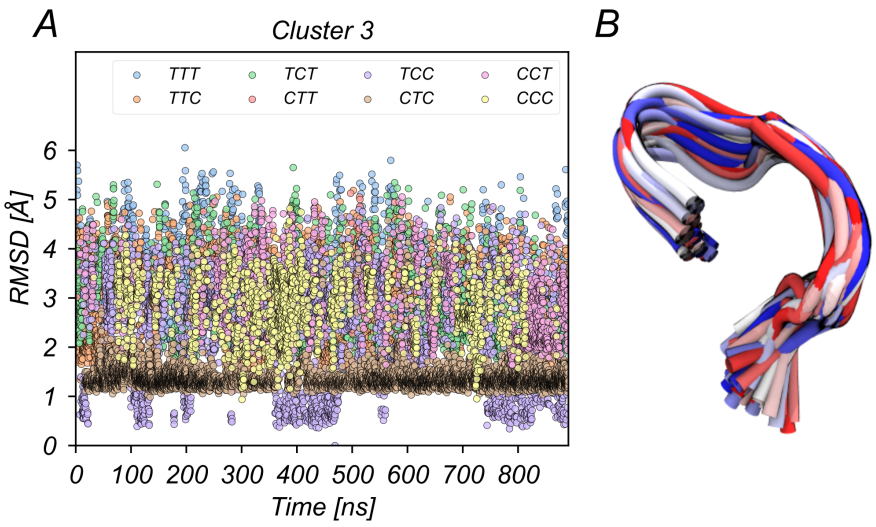



**Fig. B.3.** Characterization of the eight different isomerization states of bradykinin (BK) with respect to sampling in cluster 3 and illustration of structures sampled in cluster 3. The underlying clustering was done in a combined fashion on the entire trajectory of replica 1. (A) Root mean square deviation of the BK $C_\alpha$ atoms with respect to the configuration closest to the centroid of cluster 3, evaluated after a rototranslational fit to these atoms. The data is depicted separately for the eight BK isomers TTT, TTC, TCT, CTT, TCC, CTC, CCT and CCC. Here, the first, second and third letters refer to the trans ("T") or cis ("C") states of Pro2, Pro3, and Pro7, respectively. The trajectories pertaining to these isomerization states were filtered out from replica 1 of the $\omega$BP-REMD simulation and contained 3236, 3687, 1845, 92, 3565, 6857, 1353 and 1642 frames, respectively. Each dot represents a simulation frame. (B) Backbone superposition of the members of cluster 3. For clarity of visual illustration, only every $5^{th}$ frame along the trajectory containing the members of this cluster is shown. Colorcoding from blue to red indicates the timestep of the trajectory (dark blue, white and dark red indicating the beginning, middle and end of the trajectory, respectively).

| acceptor atom | donor hydrogen atom | donor atom | fraction [%] |
|---|---|---|---|
| | state CCT | | |
| SER6@OG | PHE8@H | PHE8@N | 44.7 |
| | state CTC | | |
| PRO2@O | PHE5@H | PHE5@N | 82.6 |
| ARG9@O | ARG1@HE | ARG1@NE | 37.4 |
| ARG9@OXT | ARG1@HE | ARG1@NE | 36.1 |
| ARG9@OXT | ARG1@HH21 | ARG1@NH2 | 31.4 |
| ARG9@O | ARG1@HH21 | ARG1@NH2 | 27.7 |
| | state CTT | | |
| PRO2@O | PHE5@H | PHE5@N | 48.9 |
| SER6@OG | PHE8@H | PHE8@N | 35.9 |
| | state TCT | | |
| SER6@OG | PHE8@H | PHE8@N | 34.9 |
| ARG9@OXT | ARG1@HE | ARG1@NE | 28.1 |
| | state TTT | | |
| SER6@OG | PHE8@H | PHE8@N | 56.8 |
| PRO3@O | SER6@H | SER6@N | 40.1 |
| SER6@O | ARG9@H | ARG9@N | 30.8 |
| PRO2@O | ARG9@HH11 | ARG9@NH1 | 30.3 |
| ARG9@OXT | ARG1@HE | ARG1@NE | 28.8 |
| ARG9@OXT | ARG1@HH21 | ARG1@NH2 | 28.2 |
| | cluster 1 | | |
| PRO2@O | PHE5@H | PHE5@N | 84.7 |
| ARG9@O | ARG1@HE | ARG1@NE | 38.3 |
| ARG9@OXT | ARG1@HE | ARG1@NE | 37.3 |
| ARG9@OXT | ARG1@HH21 | ARG1@NH2 | 31.9 |
| ARG9@O | ARG1@HH21 | ARG1@NH2 | 28.8 |
| | cluster 2 | | |
| PRO3@O | SER6@H | SER6@N | 94.6 |
| SER6@OG | PHE8@H | PHE8@N | 78.6 |
| PRO2@O | ARG9@HH11 | ARG9@NH1 | 74.1 |
| ARG9@OXT | ARG1@HE | ARG1@NE | 67.9 |
| ARG9@OXT | ARG1@HH21 | ARG1@NH2 | 66.2 |
| SER6@O | ARG9@H | ARG9@N | 55.0 |
| | cluster 3 | | |
| ARG1@O | PHE5@H | PHE5@N | 86.4 |
| SER6@O | ARG1@HE | ARG1@NE | 42.4 |
| PHE5@O | ARG1@H2 | ARG1@N | 36.2 |
| PHE5@O | ARG1@H1 | ARG1@N | 34.1 |

**Tab. B.1.** Hydrogen bonds present in BK isomers CCT, CTC, CTT, TCT and TTT, or in clusters 1, 2 and 3. In the BK state specification, the first, second and third letters refer to the trans ("T") or cis ("C") states of Pro2, Pro3, and Pro7, respectively. Here, hydrogen bonds were defined via a donor-acceptor distance of less than 3.5 Å and a donor-hydrogen-acceptor angle of greater than 135°. Only hydrogen bonds with an occurrence in at least 25% of the respective simulation frames are listed and the corresponding fraction of occurrence is reported. For BK states CCC, TCC and TTC, no hydrogen bonds were found with a minimum occurrence of 25%. Atom name specifications refer to the nomenclature used in the AMBER topology.

| state | VDW energy [kcal·mol$^{-1}$] |
|---|---|
| TTT | $-0.2 \pm 0.01$ |
| CTT | $-0.2 \pm 0.02$ |
| CTC | $-2.0 \pm 0.02$ |
| CCC | $-1.9 \pm 0.07$ |
| CCT | $-0.2 \pm 0.02$ |
| TCC | $-1.7 \pm 0.06$ |
| TCT | $-0.2 \pm 0.02$ |
| TTC | $-1.8 \pm 0.02$ |

**Tab. B.2.** Average van-der-Waals (VDW) interaction energies between residues Phe5 and Phe8 of BK in the eight possible BK isomer states. In the BK state specification, the first, second and third letters refer to the trans ("T") or cis ("C") states of Pro2, Pro3, and Pro7, respectively. Statistical errors are calculated based on block averaging. [288]

# C

# Supplementary Information for Chapter 5

| Substrate Binding Domain | Apo Conformation | Holo Conformation |
|---|---|---|
| SBD2(Fulyani et al.) | 4KR5 | 4KQP |
| OpuAC (Wolters et al.) | 3L6G | 3L6H |
| PhnD (Alicea et al.) | 3S4U | 3QUJ |
| MalE (Sharff et al.) | 1OMP | 1ANF |
| FbpA (Shouldice et al.) | 1SI1 | 1SIQ |

**Tab. C.1.** Crystal structures of several substrate binding domains, sharing a common D1/D2 compositon, connected via antiparallel $\beta$-sheets. All of these SBD's form increased C-tail contacts to the D2 domain their the open (apo) conformations.

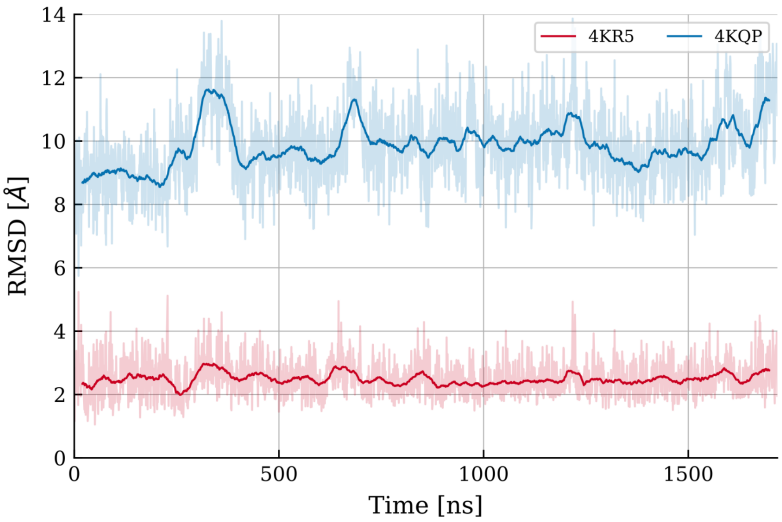



**Fig. C.1.** RMSD of unliganded open conformation of SBD2 with respect to the crystal structures of open (PDB: 4KR5, red) and closed (PDB: 4KQP, blue) conformation. In the absence of L-glutamine (GLN) ligands the RMSD stayed in the range of $\sim$ Å within over 1.8 $\mu s$ simulation time and no indication of spontaneous global domain closing.

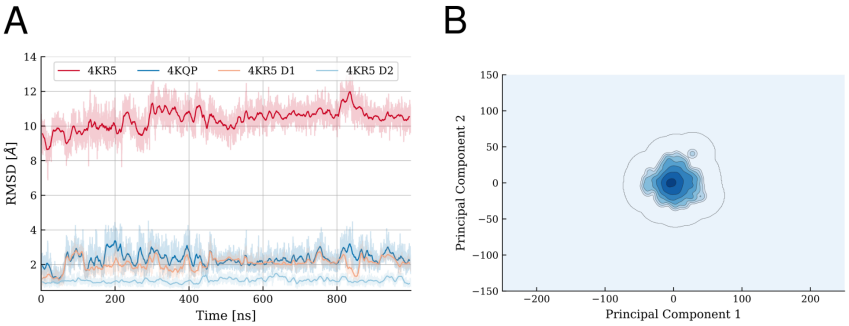



**Fig. C.2.** Free simulation of the closed, liganded SBD2. (A) Sampled RMSD values with respect to the crystal structures (open state: red; closed state: blue) and the two domains of the open conformation (D1: orange; D2: lightblue). (B) The same trajectory projected on to the first two eigenvectors. The protein remains closed, showing no major deviations from the crystal structure (Pdb: 4KQP)

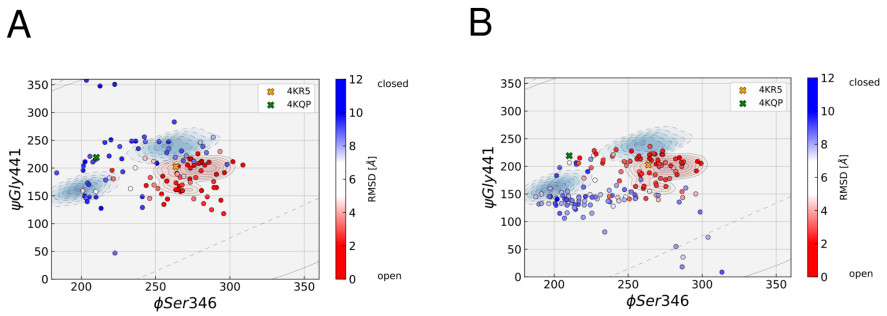



**Fig. C.3.** Conformational changes in the $\beta$-sheet hinge region upon the opening/closing of the L480A mutated SBD2. Distinct $\phi$-SER346 and $\psi$-Gly441 angle distributions along the different transitions (A) Closing transition with the ligand GLN present in the binding pocket. (B) Both closing as well as opening in the simulation without ligand (see also Figure ... )

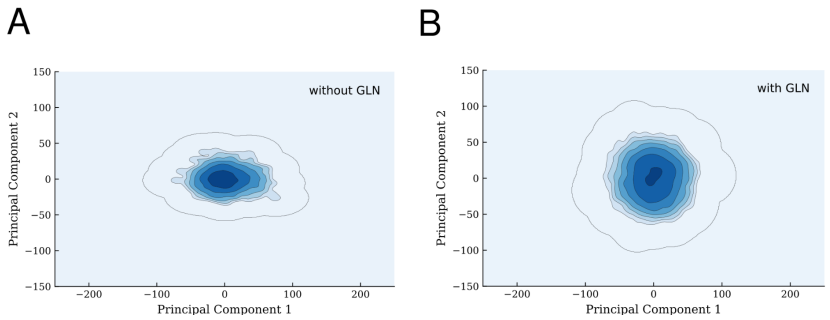



**Fig. C.4.** Principal Component Analysis of extensive unrestrained MD simulations of the open SBD2 state (PDB: 4KR5). (A) PCA of the open-unliganded SBD2 simulation (see Figure S2). (B) Addition of SBD2's specific ligand GLN does not trigger a global transition in the simulated timescales to the closed state (see also RMSD plots shown in Figure ... ).

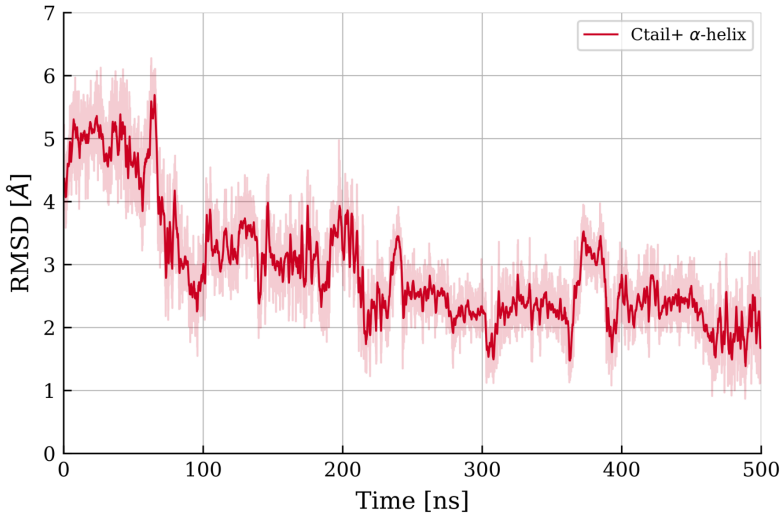



**Fig. C.5.** RMSD of C-terminal D1-$helix_{471-484}$ – D2-$helix_{418-427}$ regions with respect to the crystal structure of SBD2's open conformation upon the global opening transition (see Figure 4).

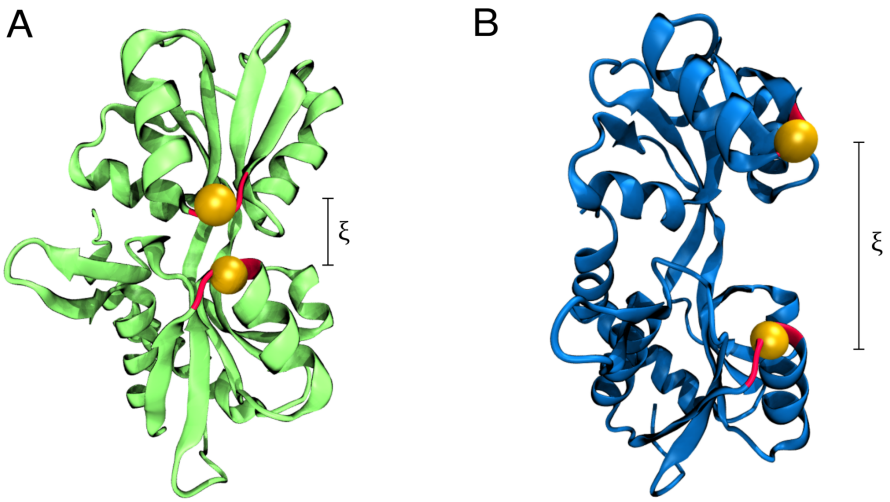



**Fig. C.6.** Illustration of the center-of-mass (yellow spheres) reaction coordinate $\xi$ (backbone atoms of residues 306-308 of the D1 and residues 396 and 397 in D2, marked in red) for H-REUS simulations of the the opening/closing transition of SBD2. (A) The crystallized structure of the closed ligand bound SBD2 (PDB: 4KQP) corresponds to a value of 7.7 Å $\sim 7.7$Å for the distance $\xi$ between the two centers. (B) In the open conformation (PDB: 4KR5) $\xi$ reaches values of $\sim 20$Å.

# D

# Supplementary Information for Chapter 6

A

| Center | COM Gln | COM Arg |
|---|---|---|
| P1 | :326@CA,C,N,O,:270@CD1,CE1,CZ,CE2,CD2,CG,:443@CD1,CE1,CZ,CE2,CD2,CG | :326@CA,C,N,O,:270@CD1,CE1,CZ,CE2,CD2,CG,:443@CD1,CE1,CZ,CE2,CD2,CG |
| P2 | :261-266,299-305@CA,C,N,O | :261-266,299-305@CA,C,N,O |
| P3 | :369-372,412-415@CA,C,N,O | :369-372,412-415@CA,C,N,O |
| L1 | N,H1,H2,H3 | N,H1,H2,H3 |
| L2 | NE2,HE22,HE21 | NH1,HH11,HH12 |
| L3 | C,OXT,O | C,OXT,O |

B

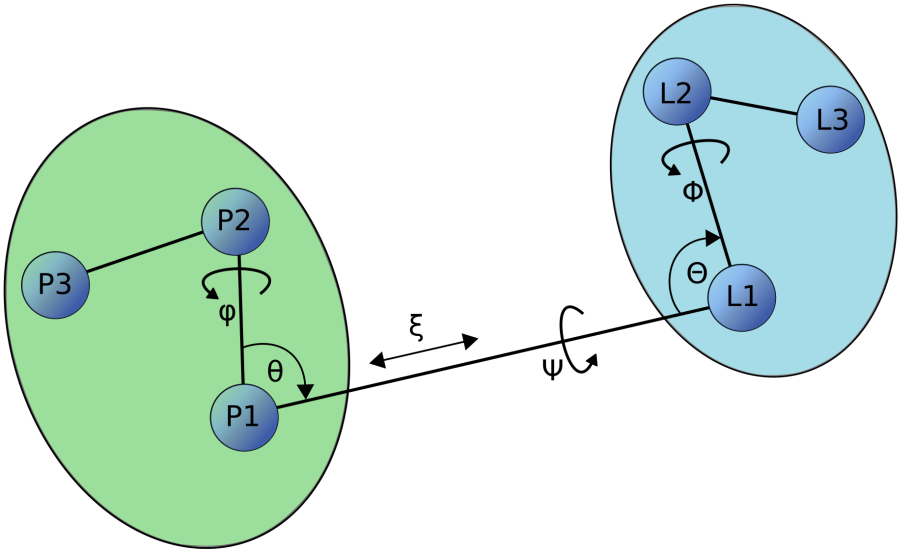



**Fig. D.1.** Supplementary Information for advanced sampling scheme . A) Center of mass definitions. Atoms given in AMBER nomenclature. B) Schematic depiction of the local reference frame used to define the position and orientation of the ligand relative to the receptor. The spherical coordinate system establishing the position of the ligand relative to the protein are the P1–L1 distance $\xi$, the P2–P1–L1 angle $\theta$, and the P3–P2–P1–L1 dihedral angle $\phi$. The Euler angles needed to define the orientation of the ligand relative to the protein are the P1–L1–L2 angle $\Theta$, the P2–P1–L1–L2 dihedral angle $\Psi$, and the P1–L1–L2–L3 dihedral angle $\Phi$

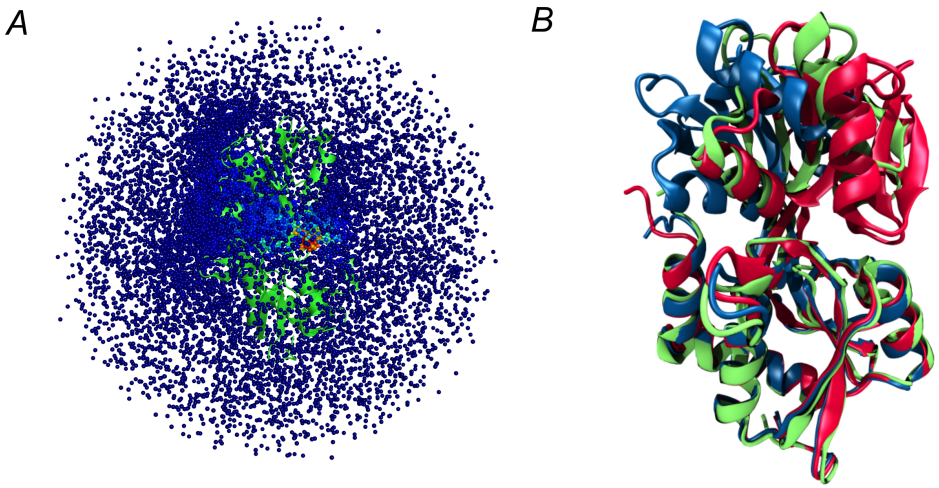



**Fig. D.2.** Supplementary Information for arginine-binding mechanism . A) Positions sampled by the ARG ligand (shown as dots, colored by density) during simulation. B) Semi-open conformations (shown in green) sampled during free simulation superimposed to the crystal structures (4KR5 in blue, 4KQP in red) . Exemplary snapshot shown.

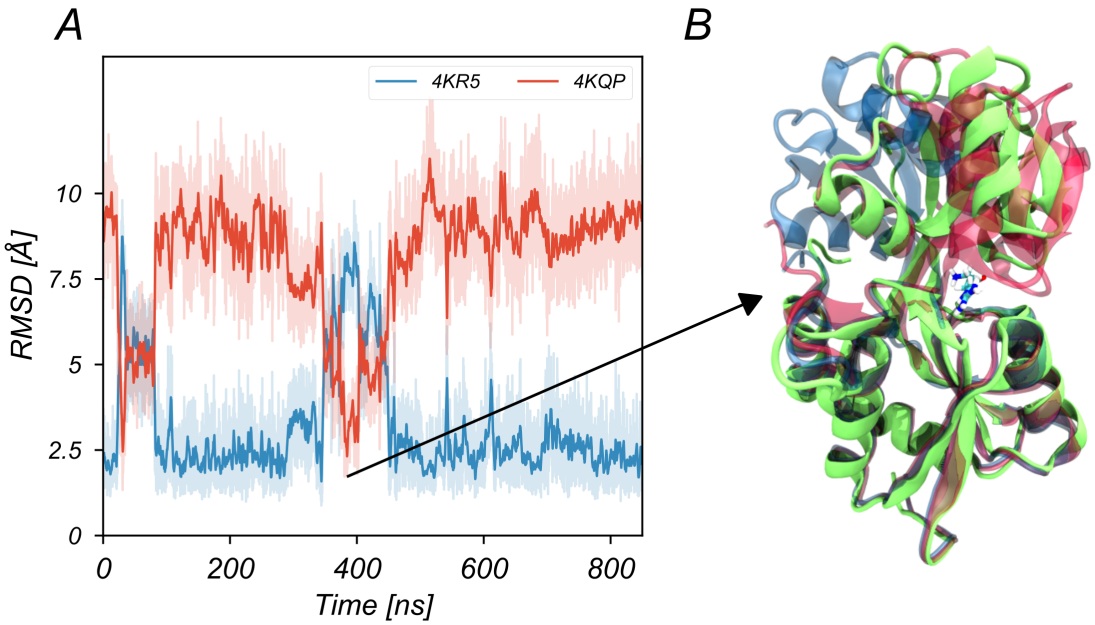



**Fig. D.3.** Arg bound to the open conformation of SBD2(L480A). A) RMSD of sampled SBD2(L480A) structures to the crystal structures of the open (in blue) and closed (red) conformations of SBD2 afer best superposition of the D1 domains. The protein never full adopts to closed state. However, there are several short transitions to a semi-open state B) Snapshot of a semi-closed frame. SBD2(L480A) is shown in green cartoon representation. The crystal structures of the open form (PDB: 4KR5) is shown in blue and the closed form (PDB: 4KQP) in red