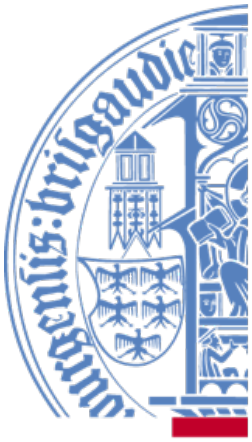


Energy and Signal Transport in Photoswitched Proteins

DISSERTATION

zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

Albert-Ludwigs-Universität Freiburg im Breisgau

vorgelegt von

Adnan Gulzar

Vorsitzender des Promotionsausschusses: Prof. Dr. Stefan Weber

Dekan: Prof. Dr. Oliver Einsle

Referent/in: Prof.Dr. Thorsten Koslowsk

Korreferent/in: Prof. Dr. Gerhard Stock

Datum der mündlichen Prüfung:

In the Name of ALLAH, the Entirely Merciful,
the Especially Merciful

Publications list

Real-time observation of ligand-induced allosteric transitions in a PDZ domain
Olga Bozovic,[a,1] Claudio Zanobini,[a,1] Adnan Gulzar,[b,1] Brankica Jankovic,[a] David Buhrke,[a] Matthias Post,[b] Steffen Wolf,[b] Gerhard Stock,[b,] and Peter Hamm[a,*]*
Submitted

Master equation model to predict energy transport pathways in proteins
Luis Valiño Borau, Adnan Gulzar and Gerhard Stock
J. Chem. Phys. 152, 045103 (2020)

Energy transport pathways in proteins:
A nonequilibrium molecular dynamics simulation study
Adnan Gulzar, Luis Valiño Borau, Sebastian Buchenberg, Steffen Wolf and Gerhard Stock
J. Chem. Theory Comput. 2019, 15, 10, 5750–5757

Photocontrolling Protein-Peptide Interactions: From Minimal Perturbation to Complete Unbinding
Brankica Jankovic,[a,1] Adnan Gulzar,[b,1] Claudio Zanobini,[a] Olga Bozovic,[a] Steffen Wolf,[b] Gerhard Stock[b,] and Peter Hamm[a,*]*
J. Am. Chem. Soc. 2019, 141, 10702-10710

Azidohomoalanine: A Minimally Invasive, Versatile, and Sensitive Infrared Label in Proteins To Study Ligand Binding
Claudio Zanobini,[a] Olga Bozovic,[a] Brankica Jankovic,[a] Klemens L. Koziol,[a] Philip J. M. Johnson,[a] Peter Hamm,[a,] Adnan Gulzar,[b] Steffen Wolf,[b] and Gerhard Stock[b,*]*
J. Phys. Chem. B 2018, 122, 10118-10125

2D-IR Spectroscopy of an AHA Labelled Photoswitchable PDZ2 Domain
Brigitte Stucki-Buchli,[a] Philip J. M. Johnson,[a] Olga Bozovic,[a] Claudio Zanobini,[a] Klemens L. Koziol,[a] Peter Hamm,[a,] Adnan Gulzar,[b] Steffen Wolf,[b] Sebastian Buchenberg[b] and Gerhard Stock[b,*]*
J. Phys. Chem. A 2017, 121, 9435-9445

[a]*Department of Chemistry, University of Zurich, Zurich, Switzerland*
[b]*Biomolecular Dynamics, Institute of Physics, Albert Ludwigs University, Freiburg, Germany*
[1]*authors contributed equally*
[*]*corresponding authors: peter.hamm@chem.uzh.ch and stock@physik.uni-freiburg.de*

Abstract

Energy and Signal transport in photoswitched proteins

Structural dynamics in proteins forms the basis of phenomena such as vibrational energy transport, protein-peptide interactions and allosteric mechanism in proteins. To obtain a detailed understanding of these processes, we prepare the systems in a nonequilibrium state via *T-jump* or *Photo-switching* methods. By employing a joint computational-experimental strategy we then follow the relaxation of the systems towards equilibrium. To that end, we first developed and employed efficient simulation protocols that mimic the experimental procedures for the investigation of the aforementioned processes. With regards to vibrational energy transport, extensive nonequilibrium energy transport simulations quantitatively reproduce the experimentally found cooling times of the proteins at room temperature and predict that the cooling slows by a factor 2 below the glass temperature of water. We further investigate the energy transport pathways and their link with allosteric communication by employing PDZ3 domain as a model system. Using these nonequilibrium simulations, we parameterized the master equation model and determined two scaling rules to predict the rates and pathways of energy transport in general proteins. The master equation model reveals that energy transport is of diffusive nature and the peptide chain along the backbone provides the fastest channel for energy transport with transfer times of 0.5 - 1 ps between adjacent residues. Contact transport, on the other hand, is considerably slower (6 - 30 ps) at room temperature. Furthermore, a quantum correction of factor ≈ 3 of the master equation parameters reproduces the experimental timescales. Second, we have considered PDZ2 domains and demonstrated that it is a minimally invasive, versatile, and sensitive infrared label that can be used to study energy transport, protein-peptide interactions and allostery because the azido ($-N_3$) group of Aha label not only reports on large changes of its chemical environment, but also on small changes of the electronic/electrostatic environment. Third, we have modelled a photoswitchable peptide which is designed in such a way that peptide-protein binding can be controlled by modulating the helical content of peptide. A direct comparison of the simulated α-helical content and experimentally obtained dissociation constants reveals that the binding affinity of S-peptide to S-protein in the Rnase S complex is inversely related to the helicity of S-peptide. Finally, using time-resolved vibrational spectroscopy, nonequilibrium molecular dynamics simulations, and subsequent Markov modeling, we describe the allosteric response of ligand-switched PDZ2 domain as a change in rugged free energy landscape, with a few structurally well defined states and the dynamics distributed over four decades starting from a nanosecond to microseconds timescales.

Acknowledgements

First of all, I am grateful to ALLAH Almighty for blessing me with the steadfastness, a wonderful research team and the strength to complete this thesis.
I wish to express my sincere gratitude to Prof. Gerhard Stock for supporting and guiding me during my doctoral research. Without Prof. Stock's continuous help and encouragement, the goals of this work would not have been realized. I would also like to extend my kind regards to Dr. Steffen Wolf for always providing me with feedback and showing me the right directions during the times of confusion. I want to specially thank Prof. Thorsten Kosliowski for providing me with doctoral admission in the Faculty of Chemistry and Pharmacy. I would like to recognize the invaluable assistance of Sebastian Buchenberg, Matthias Ernst, Mithun Biswas and Florian Sittel during the first phase of my doctoral studies.
I would like to thank Luis Valiño for all the technical help during this work and especially for the mensa discussions about energy transport, science and society, and Matthias Post for maintaining a delightful environment in the office and fruitful discussions about our research. I am also thankful to Marion Furtwängler-Fritz for making additional efforts to help me with the paperwork. I wish to thank all the team members, Benjamin Lickert, Anna Weber, Daniel Nagel, George Diez and Miriam Jaeger whose presence and assistance was a milestone in the completion of this work. I offer special thanks to Brahim Bessif, Yaser AlShetwi and Mustapha Mrabti for their thoughtful advises and encouragements during the times of despondence.
I want to offer my heartfelt thanks to all the people including Prof. Gerhard Stock, Marion Furtwängler-Fritz, Dr. Steffen Wolf, Asad Jamal, Luis Valiño, Mitthun Biswas and Matthias Ernst who supported and helped me to get back on my feet after i suffered sever injuries due to an almost life-taking accident. I am especially indebted to the hospital staff, nurses and doctors who spent their days and nights to treat me and took care of me in the dire hour of my life.
Finally, I wish to acknowledge the great moral support and unconditional love of my family throughout these years and the life.

Contents

Chapter 1

Introduction

Proteins and nucleic acids are the two main classes of biomolecules essential to all known forms of life.[1] The segments of *deoxyribonucleic acid* (DNA) known as genes are transcribed into *ribonucleic acid* (RNA), and RNA is then translated into proteins.[2] The resulting proteins then fold into three-dimensional configurations and interact with other proteins, peptides, nucleic acids, small molecules, or ions to design finely-tuned tools for carrying out specific cellular functions.[3] The ability of cells to carefully regulate transcription, translation, protein folding, modification, and function[4] is indeed a perfect order of biological life. Three dimensional folds of proteins are the end products of the information pathway that starts with the transcription of DNA. In fact, each gene in cellular DNA contains the code for a unique protein structure. Most remarkably, the coding process generates proteins with strikingly different properties and activities by covalently linking the same ubiquitous set of 20 amino acids in many different combinations and characteristic linear sequences.[5] Due to the unique chemical properties of the side chains, the specific sequence of these 20 precursor molecules may be regarded as the language of protein structure.[6] From these building blocks different organisms can make such widely diverse products as enzymes, hormones, antibodies, transporters, muscle fibers, and a number of other substances having distinct biological functions.[7]

Traditionally, the function of a protein has been thought to be completely dependent on its specific three-dimensional structure. However, recent studies have revealed that it is the synergy between structure and dynamics that determines their biological function.[8–10] Structural dynamics involve both equilibrium processes, such as thermally driven fluctuations,[11] as well as nonequilibrium processes, such as the conformational transition in a light-triggered protein.[12] These dynamic processes usually span over a large hierarchy of timescales from fast fluctuations of individual atoms and bond vibrations on femtosecond-picosecond timescales, loop and domain motions on the nanosecond timescale and conformational rearrangements on the micro- to millisecond timescales.[9] Complete understanding of the dynamical and structural properties of proteins on multiple timescales requires the description of conformational states, their transition rates, and high-resolution structures. Although state-of-the-art experimental techniques coupled with sophisticated biophysical methods can be employed to infer dynamical properties such as conformational states and respective transition rates, an atomic-resolved

dynamical description is not straightforward in these experiments. Molecular dynamics (MD) simulations, on the other hand, can describe protein dynamics in high detail: the precise position of each atom at any instant in time can be followed, along with the corresponding velocities and energies provided that at least one high-resolution structure is known as a starting point and an accurate description of the protein–solvent system could be provided by the force fields (functional form and parameter sets describing the potential energy of the system).[13]

Allostery represents an essential mechanism that regulates intra- and inter-protein signal transport. The (un)binding of signalling molecules such as ligands to the proteins or photo-activation of particular chromophores in proteins may trigger conformational or dynamical events that couple two distant sites.[14–25] Commonly, allosteric communication has been described as a series of structural changes[19–22] or shifts in conformational state populations.[23–25] It has also been suggested that allostery may be of dynamic nature i.e protein response to ligand arises from changes in thermal fluctuations without any significant structural change.[26–28] Similarly, simultaneous structural and dynamical changes were also reported as possible mechanism of allostery.[29] In any case, the nonequilibrium dynamical or structural process governing allosteric communication has not been thoroughly investigated. In addition, numerous studies relate energy transport pathways with the pathways of allostery.[30–33] It is clear that vibrational energy transport and allostery are two distinct physical processes, well separated in the hierarchy of timescales.[15] Nevertheless, the mechanical coupling connecting allosteric sites may coincide with pathways of effective vibrational energy flow.

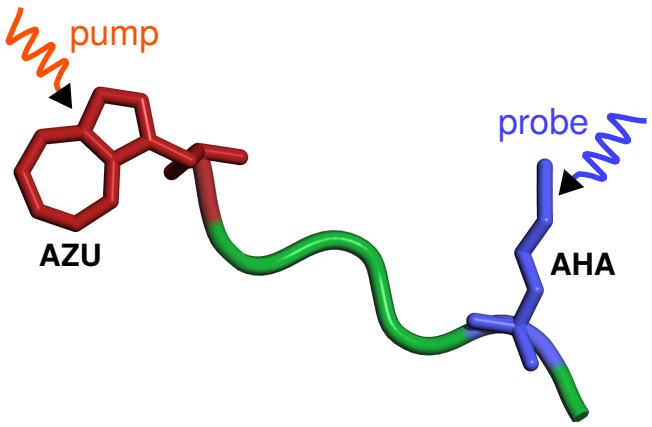


FIGURE 1.1: Scheme of donor-accepter pair employed in pump-probe spectroscopy. Vibrational energy is injected into the system via Azu (red sticks) and probed at Aha (blue sticks).

In collaboration with the Hamm lab[34] and the Bredendeck group[35] we have employed a joint experimental-theoretical approach in this work to study vibrational energy transport and signal transport. To gain a detailed understanding of the mechanisms, structural and dynamical aspects of these phenomena require to firstly prepare the system

in a well-defined nonequilibrium state and then monitor the time evolution towards equilibrium.[36] To that end, the Bredendeck group employs ultrafast pump-probe spectroscopy to inject vibrational energy into the proteins/peptides via an azulene (AZU) chromophore (serves as "heater") which can be excited at 600 nm. The azido (-N_3) group of azidohomoalanine (AHA) then serves as "probe" which exhibits temperature sensitive vibrations in Infrared (IR) window at 2100 cm^{-1} that can be used to track the vibrational energy transport in the system (as an illustration see Fig 1.1).[37] In MD simulations, we inject the vibrational energy into the system via locally heating the side chain of the heater. Then the system is allowed to relax from the nonequilibrium initial condition towards equilibrium and the propagation of vibrational energy can be monitored in space and time.[38] Nonequilibrium energy transport simulations have an additional advantage over experiments in that it is not strictly required to introduce special labels for exciting and probing the vibrational energy transport. Energy content of any unit of the system at any point in time can be measured directly. However, to make a direct comparison with the experiments of Bredenbeck and coworkers[39] we prepare the system according to experimental initial conditions i.e, we use the sidechain of azulene (AZU) as a heating moiety and attach AHA at experimentally prescribed positions. The simulation part of chapter 2 outlines the detailed strategy to mimic the experimental conditions.

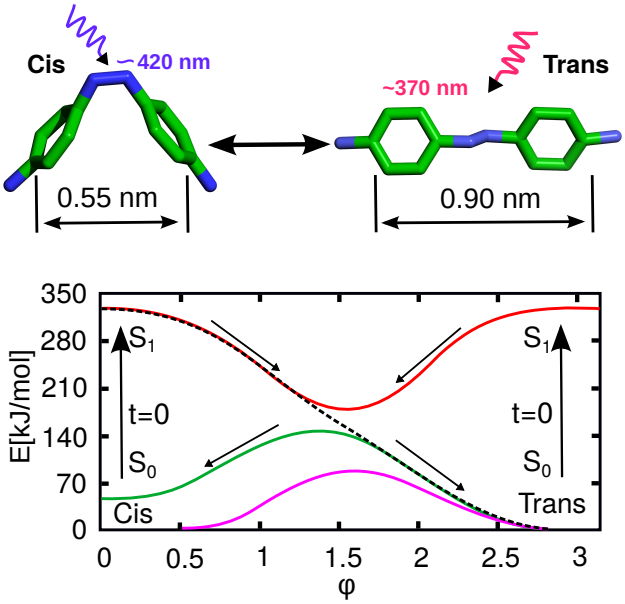


FIGURE 1.2: (left) Cis (right) Trans isomers of the azobenzene photoswitch. (bottom) N=N torsional potential of the photoswitch. (green) the S_0 ground state, (red) the S_1excited state, (pink) default parametrization in the GROMOS force-field.[40] At time, t=0, the N=N torsional potential is instantly switched from ground state S_0 to excited state S_1. The system isomerizes along excited-state N=N torsional potential within $\approx$ 250 fs (dashed lines) and excess energy is redistributed into the system and surroundings within tens of picoseconds.[41–43]

To study allosteric signal transport, Hamm and coworkers attach an azobenzene photoswitch to a peptide ligand or protein.[44] The central diazene (N=N) double bond can be reversibly photo-switched between *cis* and *trans* isomers upon illumination at different wavelengths (see Fig 1.2). In this way a nonequilibrium process is triggered on ultrafast timescales. Transient IR and 2D-IR spectroscopy is then used to probe time-resolved nonequilibrium response. To model the experimental laser-induced isomerization process of the azobenzene photoswitch in MD simulations, we employ potential-energy surface switching approach that diabatically connects the excited-state S_1 of the *cis* isomer with the ground state S_0 of the *trans* isomer and vice versa.[38,40] The isomerization leads to a local structural perturbation which travels across the proteins on multiple timescales.
To obtain the site-specific information or probe energy content, pump-probe experiments use the azide group of AHA as an IR probe. AHA is most commonly employed in vibrational spectroscopy to report on local structural changes induced upon equilibrium or nonequilibrium perturbation. By employing 2D IR spectroscopy, classical and quantum MD calculations, in this work, we further investigate the capability of AHA as a sensitive IR label.

Computational Methods

Molecular dynamics (MD) simulations are a major tool in computational studies of biomolecules.[45–47] MD simulations calculate the time dependent behavior of a molecular system thus providing detailed information about the conformational changes and dynamics of proteins with high spatial and temporal resolution. In this thesis we mainly used GROMACS software with hybrid GPU-CPU acceleration scheme to perform MD simulations.[48] Due to advances in supercomputing devices and algorithms, MD simulations softwares nowadays routinely generate microsecond to millisecond trajectories[49,50]
To effectively interpret this huge amount of simulation data, it is often required to systematically reduce the high-dimensional data to low-dimensional reaction coordinates (*collective variables*). We employ principal component analysis (PCA) for dimensionality reduction.[51] Depending upon the individual system one may choose dihedral angles, contacts or inter-atomic distances as an input high-dimensional data.[52] Based on collective variables we identify high-density clusters in low-dimensional space by using density-based clustering.[53,54] By using appropriate input parameters for clustering (e.g, lumping radius, minimum population of a cluster, and coring time) density-based clustering usually identifies metastable conformational states.[55] These metastable conformational states then serve as input for Markov state modeling (MSM). A MSM describes the dynamical process in terms of memory-less jumps between states and assists in the interpretation of experimental findings.[56]

Molecular dynamics simulation

The exact solution of the dynamics of may-particle system such as proteins requires the solution of time-dependent Schrödinger equation to the Hamiltonian operator **H** with the wave function $\psi(r,t)$

$$ih\frac{\partial}{\partial t}\psi(r,t) = \mathbf{H}\psi(r,t) \tag{1.1}$$

The numerical solution of the time-dependent Schrödinger equation, however, is prohibitive. Three main approximations are therefore required to describe the biomolecular systems in solution.[57] Due to higher masses and lower velocities of nuclei as compared to electrons, the electronic and the nuclear motions can be separated. It leads to a molecular wave function in terms of electron positions r_i' and nuclear positions r_i. This separation is known as $Born - Oppenheimer$ (BO) approximation.

$$\psi_{molecule}(r_i', r_i) = \psi_{electron}(r_i', r_i) \cdot \psi_{nulei}(r_i) \tag{1.2}$$

For a specific position of nucleus r, the Schrödinger equation can be decomposed into the time-independent Schrödinger equation for electronic wavefunction. Only the nuclear motions are to be considered, with the electronic degrees of freedom influencing the dynamics of the nuclei in the form of a *potential energy surface* (PES).
Nuclei may be considered as heavy enough to avoid quantum effects, so the second approximation replaces the time-dependent Schrödinger equation by Newton's equation of motion

$$F_i = m_i\frac{d^2r_i}{dt^2} = -\nabla_i V(r_1.....r_N) \tag{1.3}$$

with the position r_i and mass m_i of the ith nucleus.
The calculation of electronic energies for a corresponding nuclear coordinate presents another computational challenge for determining the dynamics of biomolecular systems, so the third approximation is due to the computational expense of computing the electronic potential, the ground state energy $V(r)$ is replaced by an empirical potential function known as force field. It should be noted that the motion of nuclei is interchangeably named as the motion of atoms in MD simulations. Using the simple ball-and-spring approximation,[58] force fields describe both bonded and non-bonded interactions between the atoms of the system:

$$V = V_{bonded} + V_{non-bonded} \tag{1.4}$$

Bonded interactions in force fields are described as the sum of different harmonic potential functions, assuming that the steric energy of a molecule arises from bond stretching or compressing , angle bending , bond rotations and out-of-the plane bending of planar groups of four atoms.[59,60]

$$V_{bonded} = \sum_{ij} k^b_{ij}(r_{ij} - r^0_{ij})^2 + \sum_{ijk} k^\theta_{ijk}(\theta_{ijk} - \theta^0_{ijk})^2 + \sum_{ijkl} k^\omega_{ijkl}(\omega_{ijkl} - \omega^0_{ijkl})^2 + \sum_{ijkl} k^\phi_{ijkl}[1 + \cos(\phi_{ijkl} - \phi^0_{ijkl})] \tag{1.5}$$

where r^0_{ij}, θ^0_{ijk}, ω^0_{ijkl} and ϕ^0_{ijkl} are the equilibrium values whereas k^b_{ij}, k^θ_{ijk}, k^ω_{ijkl} and k^ϕ_{ijkl} are force constants for respective four terms.
The non-bonded terms in a force field are "through-space" short (van der Waals) and long-range (electrostatic) interactions. These interactions are computed with Leonard-Jones function and Coulomb's law respectively:

$$V_{non-bonded} = \sum_{i<j} \left(\frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} + 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) \right) \tag{1.6}$$

where σ_{ij} and ϵ_{ij} are Leonard-Jones constants, q_i and q_j are the charges of atom i, j and ϵ_0 is the dielectric constant of the vacuum. These parameters are usually obtained by fitting to experimental data or from quantum mechanical calculations. The total force on each atom of the protein at time t is calculated as the vector sum of its interactions with other atoms using the equations 1.3 and respective potential functions described in equations 1.5 and 1.6. Using these forces, any finite difference technique such as Verlet,[61] Velocity verlet[62] or Leapfrog algorithms[63] can be employed to generate molecular dynamics trajectories.

Dimensionality reduction and clustering

To obtain a relevant description of high-dimensional simulation data, a transformation to low-dimensional reaction coordinates (*collective variables* (CVs)) is generally performed. In a first step, we choose high-dimensional input coordinates, e.g., Cartesian coordinates, dihedral angles, or inter-atomic distances.[52] The low-dimensional transformation upto $\lesssim 10$ has been found to preserve the essential dynamics of the system.[64,65] On this account, we employ principal component analysis (PCA) to construct a low-dimensional space of collective variables. PCA represents a linear transformation of input coordinates $\{\varphi_i\}$ that diagonalizes the covariance matrix:

$$\sigma_{ij} = \left\langle \left(\varphi_i - \left\langle \varphi_i \right\rangle\right)\left(\varphi_j - \left\langle \varphi_j \right\rangle\right) \right\rangle \tag{1.7}$$

Projecting the MD trajectories onto the eigenvectors $v^{(k)}$ of the covariance matrix results in the principal components, $V_k = v^k \varphi$, which are naturally ordered in the descending variance of the system. The resulting low-dimensional PCA space is then transformed into a metastable conformational space (state space). To that end, we employ robust density-based clustering of Sittel and Stock.[54] First of all, a local free energy estimate

for every structure of the trajectory is computed via counting all other structures inside a d-dimensional hypersphere of fixed radius R, where R equaled the lumping radius.[55] Normalization of these population counts yields densities, which give the free energy estimate:

$$F_R = -k_B T \ln \left(P_R / P_R^{max} \right) \tag{1.8}$$

where P_R is population count with in radius R.

Markov state modeling

Markov state modeling describes the dynamics of proteins as a stochastic process on a discretized state space of disjoint metastable conformations $S_1, \ldots, S_n$ with the Markov property that at a given present state, the future is independent of the past i.e conditional transition probabilities (*transition matrix*) estimated from the simulation trajectories $x(t)$ depends only upon the present state:[66]

$$T_{ij}(\tau) \equiv P(x_{t+\tau} \in S_j | x_t \in S_i) \tag{1.9}$$

The dynamics of a MSM is completely described by the transition matrix T_{ij}, containing the conditional probabilities to jump from state i to j within the lagtime τ. To estimate the transition matrix, we first generate a transition count matrix $C \in \mathbb{R}^{(n \times n)}$ by counting the number of transitions C_{ij} from state i to j. Next, we maximize the likelihood of the Markov process for the transition probability between each pair of states to convert the transition count matrix into a transition matrix T_{ij}:[67,68]

$$T_{ij} = \frac{C_{ij}}{\sum_{k=1}^{n} C_{ik}} \tag{1.10}$$

In addition to describing the conformational dynamics of proteins using data from molecular dynamics simulations, MSMs can predict both stationary and kinetic quantities on longer timescales[66] thus providing insights into the biomolecular mechanisms (such as allostery, protein folding and functional dynamics) and facilitating a direct comparison with experiments. In this work, we use the PyEMMA software package[69] for the estimation of Markov state model.

Calculation of vibrational spectral shifts

To gain insights into the molecular mechanism giving rise to the experimental spectroscopic response of Aha label, we estimate the vibrational frequency shifts $\delta\omega$ of the azido ($-N_3$) group. Based on quantum-chemical calculations, Cho and co-workers[70,71] have shown that the azido stretch mode of Aha mainly reflects the changes in the local electrostatic environment. Building on this idea, we first calculate electric fields around

each nitrogen $N^{(1)}$, $N^{(2)}$, and $N^{(3)}$ of $-N_3$ group within certain cut-off radius and then project them on molecule-fixed coordinate frame (MFCF) of $-N_3$ group.[70]

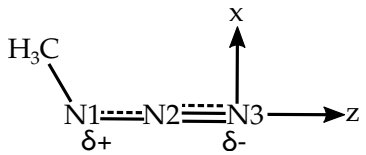


FIGURE 1.3: Scheme of molecule fixed coordinate frame of the $-N_3$ group of Aha. Difference of position coordinates of $N^{(1)}$, and $N^{(3)}$ represent z axis, and the difference of position coordinates of H_3C and $N^{(1)}$ represent vector $C\vec{N}1$.

By using above schematic 1.3 we define a cross product between given z-axis and $C\vec{N}1$, subsequent normalization then gives the molecule-fixed coordinate frame:

$$\mathbf{X}_i^{MFCF} = \begin{bmatrix} x_{x_i}, & y_{x_i}, & z_{x_i} \\ x_{y_i}, & y_{y_i}, & z_{x_i} \\ x_{z_i}, & y_{z_i}, & z_{x_i} \end{bmatrix} \tag{1.11}$$

$$\mathbf{e}_x = \frac{\mathbf{x}}{||\mathbf{x}||}, \quad \mathbf{e}_y = \frac{\mathbf{y}}{||\mathbf{y}||}, \quad \mathbf{e}_z = \frac{\mathbf{z}}{||\mathbf{z}||} \tag{1.12}$$

Finally, we project the electric fields onto a newly constructed molecule-fixed coordinate frame and calculate the Vibrational frequencies: the azido frequency shift is proportional to the electric field vector elements at three nitrogens as:

$$\delta\omega = \sum_{\alpha=x,y,z} a_\alpha^{(1)} E_\alpha^{N(1)} + a_\alpha^{(2)} E_\alpha^{N(2)} + a_\alpha^{(3)} E_\alpha^{N(3)} \tag{1.13}$$

where the nine different electric field vector components E_x, E_y, E_z at three nitrogen atoms, $N^{(1)}$, $N^{(2)}$, and $N^{(3)}$ of the azido group can be directly calculated from a given configuration of the protein. Nine Stark tuning rates $a_\alpha^{(1)}$,$a_\alpha^{(2)}$,$a_\alpha^{(3)}$ of the distributed dipoles were also taken from ref.[70]

Energy transport in proteins

This section is partly based on Gulzar et al. (2019) and Valiño et al. (2020).

In nature, many proteins undergo biochemical reactions or photoexcitation, thereby releasing heat or light energy and subsequently converting it into vibrational energy.[72] Efficient flow of vibrational energy during these processes is critical for protein structure[73,74] and function.[31,33,75–78] Moreover, the pathways of energy transport are believed to coincide with the long-range propagation of conformational change, forming the basis of allosteric communication. For example tertiary contacts such as hydrogen bonds, salt bridges and other polar contacts may provide efficient shortcuts for intra-protein energy

transport as well as play a role in the regulation of allostery.[30–33] In addition, understanding the properties of vibrational energy transport may aid in the development and improvement of molecular electronic devices.[79] Therefore, intensive experimental and theoretical studies have been carried out to elucidate the mechanism of energy transport in biomolecules.

On the experimental side, various groups employ time-resolved spectroscopy to study vibrational energy transport. By applying ultrafast vibrational spectroscopy and using local C=O vibrations as IR reporters to probe the local temperature in the proteins, Hamm and coworkers[41–43] have shown that vibrational energy diffuses along the peptide backbone on a picosecond timescale. Using a combination of time-resolved Raman spectroscopy and site-directed single tryptophan mutagenesis, the observation of energy transport with the spatial resolution of a single amino acid residue is now possible.[80,81] Alternatively, Bredenbeck and coworkers have introduced the unnatural amino acids β-(1-azulenyl)-alanine (Azu), and azidohomoalanine (Aha)[37] to track the energy flow throughout the system.

To compliment experimental findings and to provide a detailed interpretation of residue-resolved energy transport experiments[39,82] we performed nonequilibrium simulations of the anisotropic energy flow through proteins. To that end, in sections 2.1 – 2.5,[83] we describe an efficient simulation protocol that accurately mimics the excitation and probing steps of the experiments. In a first step, we considered a simple β-hairpin TrpZip2,[84] to investigate the competition between energy transport through the backbone and via various types of side chain contacts (Fig. 1.4). We then turn to the allosteric protein PDZ3, which besides its inherently complex interresidue connectivity is complicated by a conformational transition in the binding pocket region. In this system, the peptide ligand is excited, and thus the photoinduced excess energy first needs to transfer via non-covalent contacts from the ligand to the adjacent α_2 and β_2 regions of the protein (Fig. 1.4). Our next aim is to determine general quantitative scaling rules and predict energy transport pathways in proteins. Previously, Buchenberg et al.,[85,86] suggested a master equation and identified scaling rules for backbone energy transport 2.5 and via polar contacts.

$$\frac{dE_j(t)}{dt} = \sum_i \left[k_{ij}E_i(t) - k_{ji}E_j(t)\right], \tag{1.14}$$

where E_i denotes the kinetic energy of residue i and k_{ij} represents the rate of energy transport from residue i to residue j. However, Buchenberg et al. used nonequilibrium MD simulations of the small α-helical protein HP36 to parameterize the scaling rules.[86] Furthermore, simulations were conducted at low temperature (10 K) to achieve a sufficient signal-to-noise ratio, experiments are usually performed at room temperature (300K). While the master equation successfully reproduced the residue energy profiles of HP36, the applicability of the model to general proteins at experimental conditions

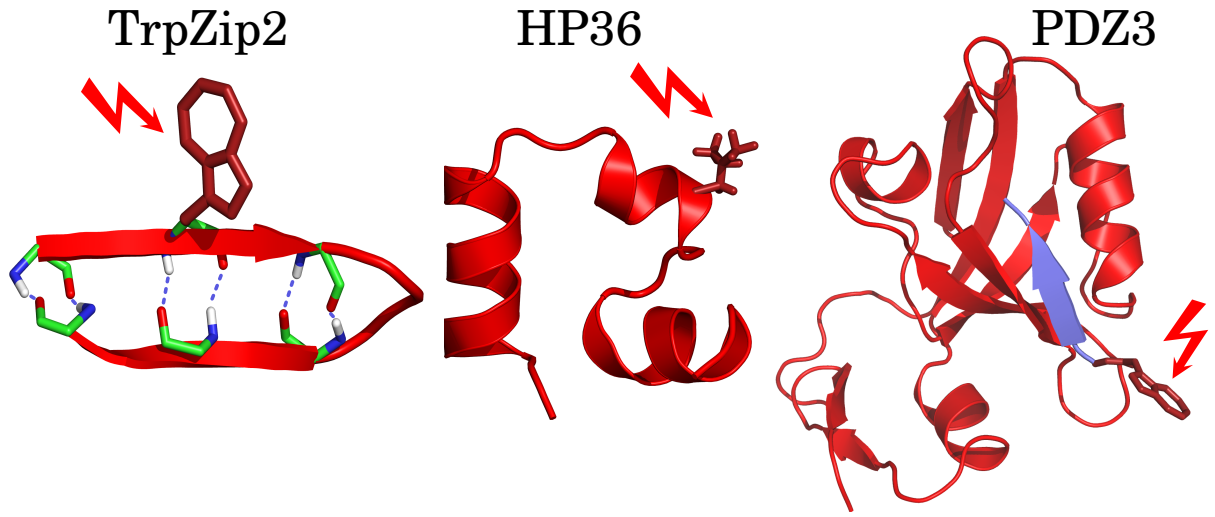


FIGURE 1.4: Scheme of the energy transport in the considered molecular systems, indicating the position of the initially excited heater residue (red), which at time $t = 0$ injects vibrational energy that is transported throughout the system via the protein backbone and interresidue contacts.

has not been investigated. To extend the master equation model to the description of energy transport in general proteins at room temperature, in section 2.6 – 2.10, we consider the nonequilibrium MD simulations of TrpZip2 and PDZ3 domain from previous sections 2.1 – 2.5 as a reference.[83] Apart from standard hydrogen bonds considered before, the modeling of these systems also requires to study the energy transport properties of aromatic contacts (e.g., formed by the Trp residues in TrpZip2), and of the contacts due to cation-π and dipole-dipole interactions formed by the Azu heater. To aid the parameterization of the model, we furthermore performed additional energy transport simulations for various model systems, including Ala12, a simplified β-hairpin model AlaZip, and HP36. In addition, we perform extensive Monte Carlo Markov chain simulations to explore the energy transport pathways within the protein.
Finally in section 2.11 we make a direct comparison between time-dependent IR experiments of Bredenbeck group,[39] nonequilibrium MD simulations and predictions of master equations by considering the residue energies of Aha in the various TrpZip2 mutations. The comparison assumes that the calculated vibrational excess energy of Aha coincides with the transient red shift ("hotband") of the azide stretch band of Aha.[87] Based on nonequilibrium energy transport simulation and subsequent fitting by the master equation, chapter 2 outlines a general theoretical approach to predict the real-time propagation of vibrational energy, rate of energy transport and pathways of energy transport in any biomolecule.

Azidohomalanine as an IR label

This section is partly based on Buchli et al. (2017) and Zanobini et al. (2018).
Vibrational spectroscopy provides an inherent picosecond time resolution to study both equilibrium and non-equilibrium processes in proteins. However, obtaining site-selective information from vibrational spectroscopy requires vibrational labels. Among other properties, such vibrational labels should absorb outside of the congested region of the absorption spectrum of proteins[88–92] to discriminate it from a substantial background. Various distinct molecular groups have been suggested for that purpose: all these molecular groups absorb in a spectral window between ≈1700 cm^{-1} and ≈2800 cm^{-1}, where no fundamental modes of natural proteins are found.

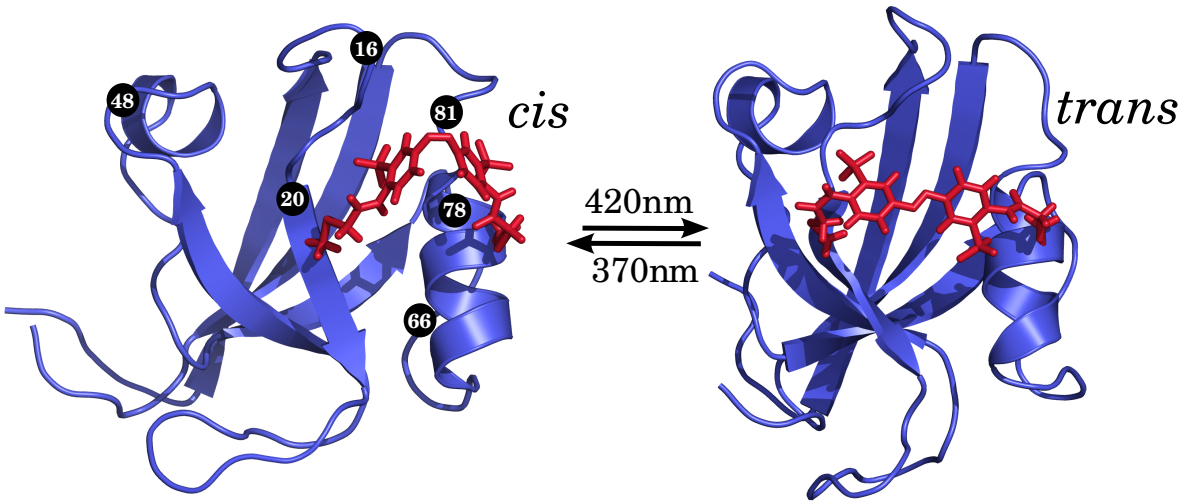


FIGURE 1.5: Photoswitchable PDZ2 domains in *cis* and *trans* conformations with the positions of amino acids mutated to Aha.

Due to its versatile properties, in chapter 3 , we focus on the non-natural amino acid azidohomoalanine (AHA) as an IR vibrational label, which contains an azido group (–N_3) that absorbs at around 2100 cm^{-1}. Our aim is to explore the capability of the AHA label to sense relatively small structural changes of a protein, i.e., changes smaller than those occurring upon folding or unfolding. To that end, in sections 3.1 – 3.5, we AHA-labelled an allosteric protein (the PDZ2 domain from human tyrosine-phosphatase 1E) and furthermore covalently linked it to an azobenzene-derived photoswitch to mimic its conformational transition upon ligand binding. To determine the strengths and limitations of the AHA label, in total six mutants have been investigated with the label at sites with varying properties. Fig. 1.5 shows all amino acids that have been mutated to AHA. We used 2D IR spectroscopy to measure the IR signal induced by photoswitching

the azobenzene moiety from *trans* to *cis* and subsequent classical simulations and quantum mechanical vibrational frequency calculations to describe the structural changes in each photoswitched state.

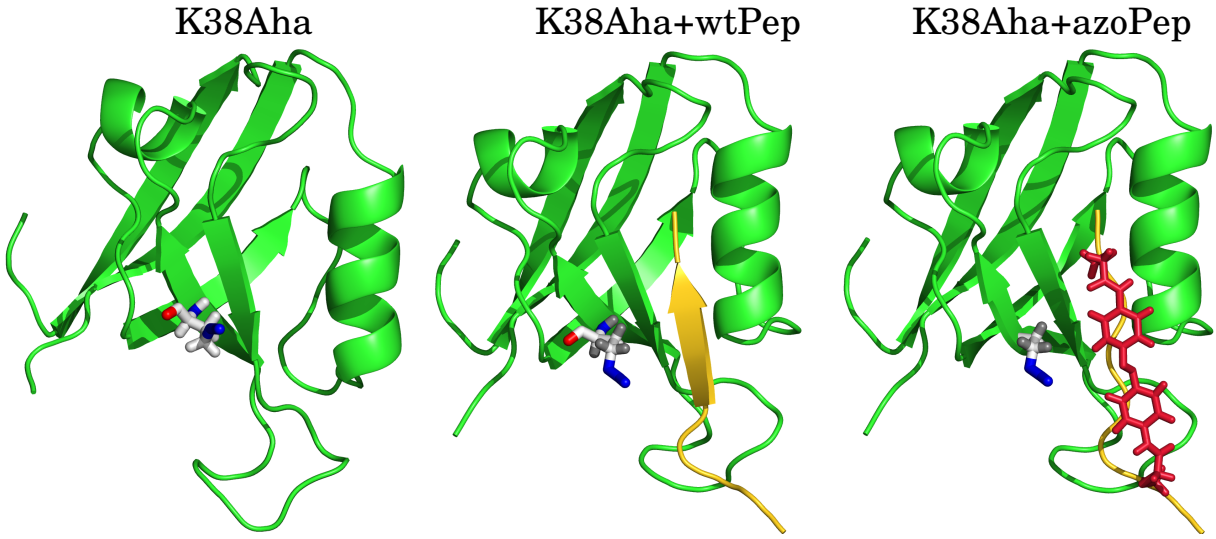


FIGURE 1.6: Molecular systems considered in the sections 3.6–3.10. (left) The K38Aha mutant of apo-PDZ2 with the Aha label indicated as sticks, (middle) with the wild-type peptide bound to its binding groove (K38Aha+wtPep), and (right) with azobenzene-variant of the peptide bound to the binding groove (K38Aha+azoPep); azobenzene photoswitch displayed as red sticks. The shown structures are snapshots taken from the MD simulation.

In recent previous works, Hamm and coworkers incorporated the Aha label into RaGEF2 peptides to investigate site-specific changes upon peptide-protein recognition.[93,94] Upon binding to a PDZ2 domain, the peptide ligand undergoes a structural change and depending upon the orientation of azido group of the Aha label, a frequency shift of 15 cm^{-1} was observed. In sections 3.6 – 3.10, we further explore the limits of this procedure. Instead of the peptide ligand,[93,94] we incorporate Aha label in the protein sequence near the binding pocket at position Lys 38 (K38). Upon ligand binding, this site was shown to be dynamically active, i.e, frequent forming salt-bridges with K38 during the process of binding .[95]

In the 3.6 – 3.10, we considered the three molecular systems shown in Fig. 1.6. As a reference, we measured the frequency shifts and simulated the K38Aha mutant of apo-PDZ2. We then bound either the wild-type peptide (RWAKSEAKENEQVSAV) to its binding groove or a mutant of that peptide (RWAKSEAKECEQVSCV) with an apolar azobenzene moiety attached to the peptide via two cysteines.

Design of photoswitchable peptides

This section is partly based on Jankovic et al. (2019).
Protein-protein/peptide interactions are a fundamental part of all major biological processes.[96] An external control of these interactions facilitates a detailed understanding of real-time conformational and dynamic changes induced by protein-peptide recognition. Photocontrol has shown to be a valuable tool for the external control of numerous biological processes. In addition to being reversible,[97] photoexcitation offers high spatial and temporal resolution and selectivity.[98] Upon illumination, the azobenzene chromophore undergoes a reversible *cis* to *trans* isomerization around its central diazene (N=N) double bond. Isomerization changes the geometry of the azobenzene moiety with different end-to-end distances of the two configurations, leading to either a perturbation or a stabilization of system.[99]

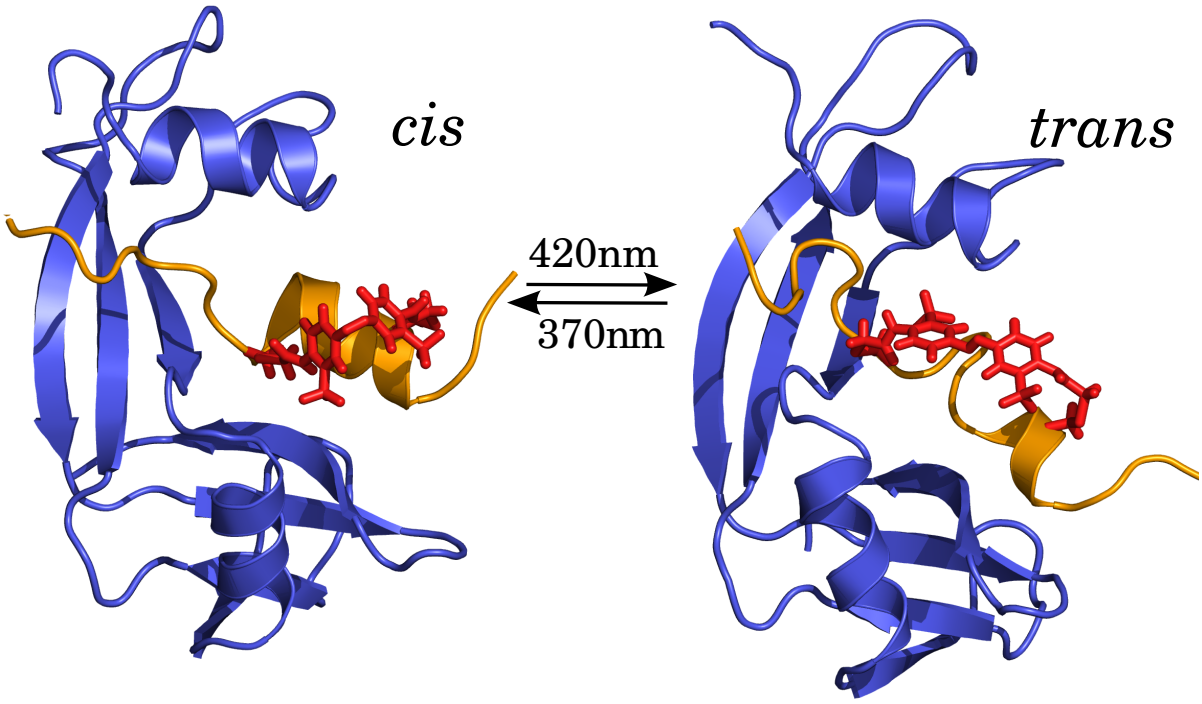


FIGURE 1.7: Molecular construct studied in chapter 4: S-protein (blue) with S-pep(6,13) (blue) bound to it in the two states of the photoswitch (orange).

Azobenzene and its various derivatives have been used to phototrigger folding and unfolding of small α-helices, study energy, and information transport in small peptides, and β-hairpins. By employing azobenzene photoswitch, we seek to design a photoswitchable peptide such that its binding affinity to the protein can be artificially modulated with the goal of achieving photoinduced peptide unbinding. In chapter

4, we demonstrate a photocontrol strategy by adopting the non-covalent ribonuclease S complex (RNase S). RNase S results from limited, site-specific hydrolysis of ribonuclease A, an enzyme from bovine pancreas.[100] Under controlled conditions, subtilisin can cleave a single peptide bond in RNase A and yields the RNase S complex composed of the S-peptide (residues 1-20) and the S-protein (residues 21-124).[100] Full enzymatic activity is restored when the two components form a native-like complex. The three-dimensional structure of the complex is same as that of intact RNase A.[101] Consequently, the S-peptide adopts an α-helical structure when it is bound to the S-protein, while it exists as a disordered peptide in isolation.[102] This property makes RNase S an excellent model to study the question whether the recognition mechanism between the S-protein and S-peptide can be characterized as "induced-fit" or as "conformational selection".[103–106] Furthermore, it opens the possibility to phototrigger binding/unbinding of the S-peptide, adopting the concept that has previously been used to control the α-helical content of isolated peptides via cross-linking two sites of the helix with an azobenzene moiety.[44,107–109] Fig. 1.7 shows the construct we designed for this work: a S-peptide (yellow) is bound to a S-protein (blue), the S-peptide is being photoswitched and designed in a way that the azobenzene-moiety (red) in its *trans*-configuration destabilizes the α-helical content of the S-peptide. To facilitate photocontrolling of peptide-protein binding, five different photoswitchable S-peptide variants have been designed.

Signal transport in proteins

This section is partly based on Bozovic et al. (2020).

Any perturbation at one site of a protein, which initiates a physical or chemical change and eventually affects a distal site, can be regarded as signal transport. For example, binding of a ligand or absorption of light may cause a conformational or dynamical change in the protein, thereby regulating the signalling pathways. Allostery is one of the most important signal transport process, vital for the regulation of various functions in all living organisms. It is generally defined as coupling of two sites in a protein or a protein complex, where the binding of a ligand to the distal site modifies the affinity at the active site.[18] Commonly, ligand (un)binding is associated with a change of the protein's mean structure.[110] On the other hand it may also alter the protein's flexibility, by changing the variance of the structure thus providing entropic contribution to the free energy.[26] In the absence of any observable conformational change, the latter scenario, is known as dynamic allostery.

In chapter 5, we consider PDZ2 and PDZ3 domains as model systems to investigate the allosteric mechanism. PDZ domains are well-known examples of dynamic allostery due to their minuscule response upon ligand (un)binding.[111–113] In sections 5.1 – 5.3, we consider a PDZ2 domain bound to a RA-GEF-2 peptide ligand[114] with an azobenzene

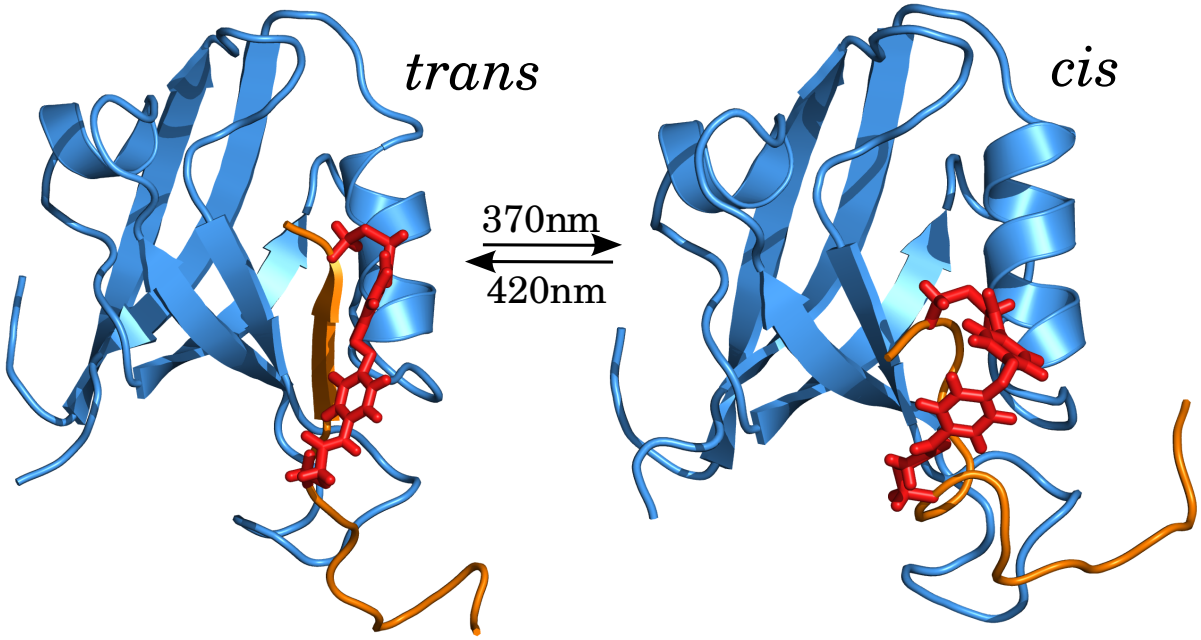


FIGURE 1.8: Ligand-switched PDZ2 domain. PDZ2 (blue) with the ligand (orange) and in *trans* and *cis*.

moiety linked as the photoswitch[115] see (Fig. 1.8). By photoisomerizing the azobenzene, we change the binding affinity of the ligand. We employ isotope-labelled time-resolved vibrational spectroscopy to monitor the structural change of the protein in real time, and perform extensive (more than 0.5 ms aggregate simulation time) all-atom non-equilibrium MD simulations combined with Markov modeling to interpret the experimental results in terms of the structural evolution of the system. We find that the mean structural change of the protein is rather small. Yet, in both experiments and MD simulations, the free energy surface of the protein can be characterized by a small number of metastable conformational states. In agreement with the view of allostery as an interconversion between the relative population of metastable states, the ligand-induced response of the PDZ2 domain is described as a remodelling of the free energy landscape,[116–120] where dynamics is distributed over four decades starting from a nanosecond to microseconds timescales.

In section 5.6 we consider the PDZ3 domain of the postsynaptic density-95 (PSD-95). In addition to the conserved fold of 5-6 β-sheets and 2 α-helices, PDZ3 contains an additional third α-helix (α_3), which packs against the core domain, located on the opposite side from the binding pocket (see figure 1.9).[121,122] Petit *et al.* showed that truncation of α_3 helix reduces the binding affinity of a CRIPT ligand by 21 fold. The enhanced flexibility of side chains was determined to be the cause of the reduction in binding affinity.[113] To mimic a similar construct, we designed a photoswitchable PDZ3 domain

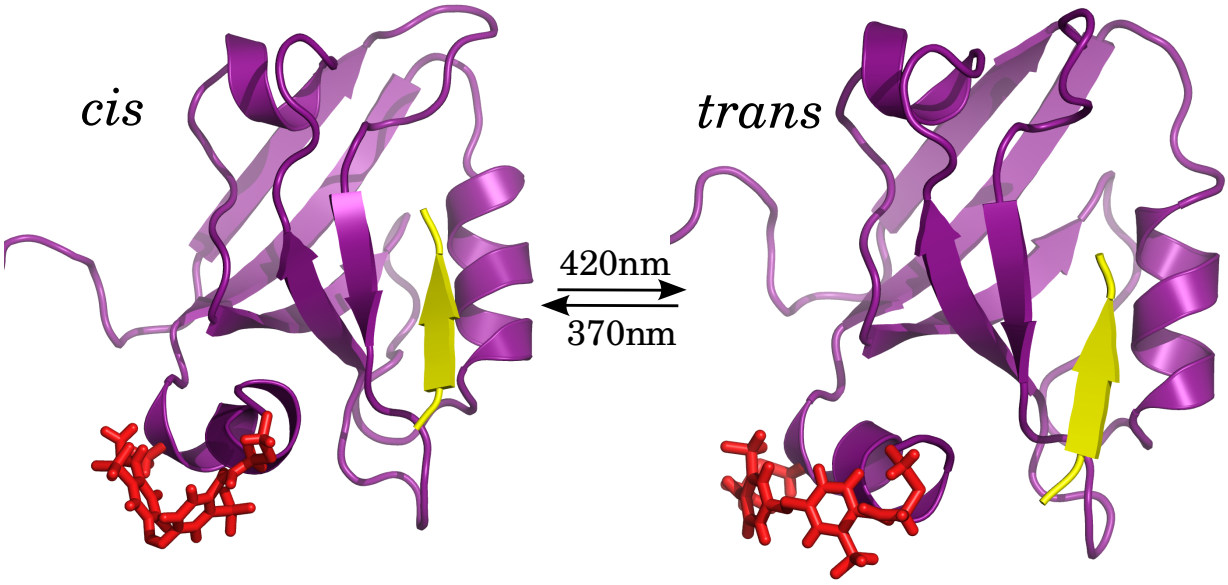


FIGURE 1.9: Cartoon structure of PDZ3 domain in purple; ligand in yellow, photoswitch in red at α_3 in *cis* and *trans* conformations.

where azobenzene-photoswitch was attached at α_3. Accompanying experiments revealed $\approx$ 25–folds reduction in binding affinity upon *cis–to–trans* photoswitching of α_3–switched PDZ3. By employing extensive nonequilibrium simulations, we aim to explore the causes of this astonishing reduction in binding affinity.

Chapter 2

Energy Transport in Biomolecules

This chapter is based on following publications:
Energy transport pathways in proteins:
A nonequilibrium molecular dynamics simulation study
Adnan Gulzar, Luis Valiño Borau, Sebastian Buchenberg, Steffen Wolf and Gerhard Stock
J. Chem. Theory Comput. 2019, 15, 10, 5750–5757

Master equation model to predict energy transport pathways in proteins
Luis Valiño Borau, Adnan Gulzar and Gerhard Stock
J. Chem. Phys. 152, 045103 (2020)

To facilitate the observation of biomolecular energy transport in real time and with single-residue resolution, recent experiments by Baumann et al. *[Angew. Chem. Int. Ed.* **58**, *2899 (2019)]* have used unnatural amino acids β-(1-azulenyl)-alanine (Azu) and azidohomoalanine (Aha) to site-specifically inject and probe vibrational energy in proteins. To aid the interpretation of such experiments, nonequilibrium molecular dynamics simulations of the anisotropic energy flow in proteins TrpZip2 and PDZ3 domain are presented. On this account, an efficient simulation protocol is established that accurately mimics the excitation and probing steps of Azu and Aha. The simulations quantitatively reproduce the experimentally found cooling times of the solvated proteins at room temperature and predict that the cooling slows down by a factor two below the glass temperature. In PDZ3, vibrational energy is shown to travel from the initially excited peptide ligand via a complex network of inter-residue contacts and backbone transport to distal regions of the protein. The energy transport pathways may coincide with pathways of allosteric communication.
As a simple means to describe these experimental and simulated data, Buchenberg et al. *[J. Phys. Chem. Lett.* **7**, *25 (2016)]* suggested a master equation model which accounts for the energy transport from an initially excited residue to some target residue. The transfer rates of the model were obtained from two scaling rules, which account for the energy transport through the backbone and via tertiary contacts, respectively, and were parameterized using simulation data of a small α-helical protein at low temperatures. To extend the applicability of the model to general proteins at room temperature, here

a new parameterization is presented that is based on extensive nonequilibrium molecular dynamics simulations of a number of model systems. With typical transfer times of 0.5 - 1 ps between adjacent residues, backbone transport represents the fastest channel of energy flow. It is well described by a diffusive-type scaling rule, which requires only an overall backbone diffusion coefficient and interatom distances as input. Contact transport, e.g., via hydrogen bonds, is considerably slower (6 - 30 ps) at room temperature. A new scaling rule depending on the inverse square contact distance is suggested, which is shown to successfully describe the energy transport in the allosteric protein PDZ3. Since both scaling rules require only the structure of the considered system, the model provides a simple and general means to predict energy transport in proteins. To identify the pathways of energy transport, Monte Carlo Markov chain simulations are performed, which highlight the competition between backbone and contact transport channels.
While nonequilibrium energy transport simulations underestimate the peak times by a factor of ≈ 2 due to the classical nature of MD simulations, a global scaling up by a factor 3.1 of master equation parameters reproduces the experimental timescales of energy transport at a single-residue resolution.

Energy transport pathways in proteins: A nonequilibrium molecular dynamics simulation study
Here, I performed all simulations and analyzed the data.

2.1 Nonequilibrium energy transport simulations

The transport of vibrational energy in biomolecules such as proteins and nucleic acids has been investigated for decades.[32,73,74,123–125] For one, this is because excess energy (e.g., due to photoexcitation or a chemical reaction) needs to be dissipated rapidly and efficiently to avoid chemical bond ruptures that would destroy the protein.[73,74] Moreover, inter-residue energy transfer has been linked with protein function.[31,33,75–78] Apart from the energy transport along the protein backbone, in particular, the transport via tertiary contacts such as hydrogen bonds, salt bridges and polar contacts is of interest. Establishing mechanical couplings between residues that are distant in sequence space, contact energy transport is believed to coincide with the long-range propagation of conformational change, which is the hallmark of allosteric communication.

Time-resolved vibrational spectroscopy has made it possible to monitor the flow of biomolecular energy in real time. For example, Hamm and coworkers[41–43] used a molecular photoswitch to impulsively heat a peptide helix and employed local C=O vibrations as read-out parameters to probe the local temperature[126] throughout the biomolecule. In a series of experiments they showed that energy transport along the peptide backbone occurs on a picosecond timescale and in a diffusive manner. Using a

combination of time-resolved Raman spectroscopy and site-directed single tryptophan mutagenesis, recently the residue-resolved observation of energy flow in heme proteins was reported.[80,81] As an alternative strategy to study the vibrational energy transfer in proteins, Bredenbeck and coworkers have introduced the unnatural amino acids β-(1-azulenyl)-alanine (Azu) and azidohomoalanine (Aha).[37] The azulene chromophore of Azu acts as a "heater" that can be excited at 600 nm and injects vibrational energy into the system via ultrafast internal conversion. The azido group of Aha acts as vibrational probe at different sequence positions, which can be monitored at 2100 cm^{-1} to track the energy flow throughout the system. Besides smaller peptides,[37] this strategy was recently applied to the third PDZ domain of the synaptic protein PSD-95 (PDZ3).[82] PDZ domains bind the C-terminus of various ligands with high specificity[127] and constitute a well-studied model for allosteric communication.[14–16,113,128–130] Specifically, Baumann et al.[82] considered PDZ3 bound to CRIPT peptide (Fig. 2.1), and inserted the Azu-Aha pair at various sequence positions.

On the theoretical side, biomolecular energy flow has been described by equilibrium and nonequilibrium molecular dynamics (MD) simulations.[31,33,75–78,130–134] Moreover, various energy transport networks have been proposed.[85,86,128,129,135–138] Only relatively few works, though, have attempted to directly predict measured signals. For one, it may not be straightforward to model the experimental excitation process (often involving excited electronic states) as well as the detection step (involving, e.g., hot-bands due to nonlinear coupling of the probed vibrational mode with thermally excited low-frequency modes[87]). In order to compare to experiment, moreover, simulations need to match the experimentally employed conditions rather than idealized ones. For example, to achieve sufficient signal-to-noise ratio, simulations have been conducted at low temperatures (such as 10 K),[75,86] while experiments are usually performed at room temperature (300 K).

Motivated by the advent of residue-resolved energy transport experiments for proteins such as PDZ3,[80,82] this work aims to extend previous MD studies on smaller systems and perform nonequilibrium simulations of the anisotropic energy flow through proteins. To this end, we first establish an efficient simulation protocol that accurately mimics the excitation and probing steps of the experiments, which also involves the force field parameterization of the Azu and Aha residues. Studying the temperature dependence of the energy transport of PDZ3, we find that at previously employed low temperatures the cooling process of the hot protein in the surrounding solvent is artificially slowed down in comparison to the situation in experiments. Hence, to compare to experiments at room temperature, simulations need to be performed at 300 K as well, which typically requires several thousands of nonequilibrium trajectories in order to achieve a sufficient signal-to-noise ratio.

In a first step, we perform a case study of the 12 amino-acid peptide TrpZip2,[84] which allows us to study in detail the competition between energy transport through

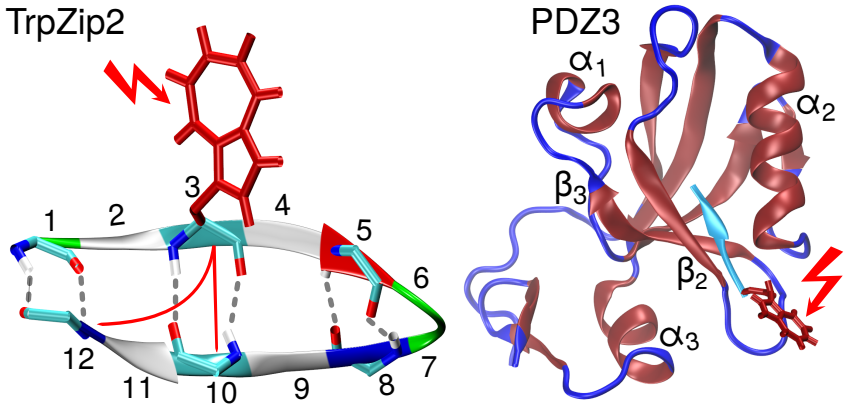


FIGURE 2.1: Scheme of the energy transport in TrpZip2 and PDZ3, indicating the position of the initially excited Azu residue (red), which at time $t = 0$ injects vibrational energy that is transported throughout the system via the protein backbone and inter-residue contacts. For TrpZip2, direct side-chain contacts with the Azu heater are indicated by red lines, and β-structure stabilizing hydrogen bond are indicated by gray lines. In the case of PDZ3, the Azu residue heats the ligand (cyan), from which the energy flows via various contacts to the protein.

the backbone and via various types of side chain contacts (Fig. 2.1). We then turn to the energy transport in the allosteric protein PDZ3, which besides its inherent complex interresidue connectivity is complicated by a conformational transition in the binding pocket region. In this system, the peptide ligand is excited, and thus the photoinduced excess energy first needs to transfer via non-covalent contacts from the ligand to the adjacent α_2 and β_2 regions of the protein (Fig. 2.1).

2.2 Computational Methods

2.2.1 Equilibrium MD simulations

All MD simulations were performed using GROMACS package v2016.3 (Ref.[48]) and the Amber99sb*ILDN forcefield.[139–141] Parameters for Azu were obtained from Antechamber[142] and Gaussian09[143] (see Supporting Information), Aha parameters were reported previously.[144] Na^+ and Cl^- were added at a salt concentration of 0.1 M, with an excess of Cl^- to compensate the net positive charge (+2) of TrpZip2, and an excess of Na^+ to compensate the net negative charge of (-3) of PDZ3. Long-range electrostatic interactions (distances > 1.2 nm) were computed by the Particle Mesh Ewald (PME) method,[145] short-range electrostatic interactions were treated explicitly using a Verlet cut-off scheme. After energy minimization, a 10 ns NPT equilibration run was performed. In all equilibrium simulations, we used an integration time step of 2 fs and maintained a temperature of 300 K (via the Bussi thermostat,[146] coupling time 0.1 ps) and a pressure of 1 bar (via the Berendsen barostat,[147] coupling time 0.1 ps).

Taking the starting structure of TrpZip2 from PDB entry 1LE1,[84] Thr3 and Thr10 were mutated to Azu and Aha, respectively. TrpZip2 was then solvated with ≈ 700 TIP3P[148] water molecules in a dodecahedron box with a minimum image distance of 3 nm. Following equilibration, a NVT production run was carried out for 100 ns. The initial structure of PDZ3 was taken from PDB entry 1BE9.[121] Azu was attached at the N-terminus of CRIPT peptide and Ile327 was mutated to Aha327. We added ca. 10 400 TIP3P water molecules in a dodecahedron box with a minimal image distance of 7.8 nm and set the PME cut-off length 1.4 nm. Following 200 ns of NVT equilibration, five 1μs long MD runs were performed.

Cooling both systems down to 100 K and 10 K, respectively, we performed similar equilibrium runs at these temperatures. From the simulations of TrpZip2 and PDZ3 at temperatures $T_0 = 10$, 100 and 300 K, N_{traj} statistically independent initial structures were stored for the subsequent nonequilibrium runs. As listed in Table 2.1, we choose the number of nonequilibrium trajectories N_{traj} from 252 for test calculations up to 10 000 for production runs.

TABLE 2.1: List of nonequilibrium simulations, including number of trajectories N_{traj}, initial solvent temperature T_0, integration time step δt, time constant of the thermostat τ_T, and cooling time τ_{C}. Error bars denote the error from the exponential fit to the MD data.

system	N_{traj}	ensemble	T_0/K	δt/fs	τ_T/ps	τ_{C}/ps
TrpZip2	2700	NVE	300	0.2	–	9.1 ± 0.018
	252	NVT	10	0.7	10	17.7 ± 0.097
	252	NVT	100	0.7	10	14.7 ± 0.054
	5000	NVT	300	0.7	10	8.8 ± 0.010
PDZ3	252	NVE	100	0.2	–	15.2 ± 0.026
	252	NVT	10	0.7	10	14.2 ± 0.055
	252	NVT	100	0.5	10	12.4 ± 0.055
	252	NVT	100	0.7	10	12.7 ± 0.050
	252	NVT	100	1.5	10	13.3 ± 0.099
	252	NVT	100	0.7	5	12.3 ± 0.075
	252	NVT	100	0.7	20	13.5 ± 0.069
	252	NVT	100	0.7	50	14.9 ± 0.060
	10000	NVT	300	0.7	10	8.3 ± 0.029

Data evaluation was carried out using Gromacs tools.[48] We define a contact to be formed if the minimal distance between the atoms of two residues is shorter than 0.45 nm.[52]

2.2.2 Nonequilibrium MD simulations

To mimic the initial heating of azulene via electronic excitation and subsequent ultrafast (∼ 1 ps) internal conversion,[149,150] we approximate the resulting vibrational excitation by an instantaneous temperature jump, where the excess energy $k_{\text{B}}\Delta T$ is chosen to match the $S_0 \to S_1$ excitation energy of ≈ 2 eV, resulting in $\Delta T \approx 600$ K. By employing the Gromacs *freeze* command, we constrain all degrees of freedom except for the

$f_H = 44$ degrees of freedom of azulene, which are coupled to a thermostat set to $T_0 + \Delta T$ and heated for 2 ps. Following the heating step, the velocities of the frozen atoms are reassigned, and nonequilibrium MD simulations of 50 - 100 ps length are carried out.

To monitor the flow of vibrational energy from the azulene heater through the protein, we consider the time evolution of the kinetic energy of the ith residue, $E_i^{\mathrm{kin}}(t) = \sum_j E_{i,j}^{\mathrm{kin}}(t)$, where the sum runs over all atoms j of residue i. The time-dependent expectation value of the kinetic energy per degree of freedom, $E_i(t)$, is calculated via an ensemble average over N_{traj} nonequilibrium trajectories,

$$E_i(t) = \frac{1}{f_i N_{\mathrm{traj}}} \sum_{n=1}^{N_{\mathrm{traj}}} E_i^{\mathrm{kin}}(n,t) - E_i^{\mathrm{eq}}, \tag{2.1}$$

where f_i denotes the degrees of freedom of residue i. Since $E_i^{\mathrm{eq}} = k_B T/2$ is the equilibrium energy per degree of freedom, $E_i(t)$ is expected to decay to zero at long times. For easier representation, all residue energies and atom group energies shown below were smoothed by a Gaussian filter. For details of the smoothing procedure, see the Supplementary Information which also compares smoothed and raw data in Fig. S1 - S4.

The total kinetic energy per degree of freedom of a protein with M residues is given by

$$E_P(t) = \frac{1}{M} \sum_{i=1}^{M} E_i(t). \tag{2.2}$$

In a similar way, the kinetic energy of the solvent per degree of freedom $E_S(t)$ and its equilibrium value E_S^{eq} are calculated. To estimate the cooling rate of the protein, a monoexponential decay behavior with cooling time τ_C is assumed

$$E_P(t) = (E_P(0) - E_P(\infty))e^{-t/\tau_C} + E_P(\infty). \tag{2.3}$$

2.3 Results and Discussion

2.3.1 Simulation strategy

Nonequilibrium energy transport calculations using high-temperature excitation present a challenge to standard MD simulation protocols that are designed for equilibrium modeling at room temperature. Due to the inherent non-ergodicity, averages of energies have to be calculated as means over separate trajectories. To reduce the signal-to-noise level for obtaining meaningful results, numerical approximations and rounding errors must be reduced to a minimum. That is, we find that double precision floating-point representation is absolutely indispensable, since for single precision accuracy the total kinetic energy may be completely off after a few picoseconds simulation time (Fig. S5).

The simulation results may even depend on the version of a simulation package. For example, Fig. S6 shows that version v2016.3 of Gromacs delivers much more stable energy estimates than the older version v4.6.7.

Special care also needs to be taken concerning the choice of the thermodynamic ensemble (e.g., microcanonical NVE or canonical NVT) and the thermostat parameters, as well as the integration time step δt. To simulate energy transport, commonly NVE conditions are employed which avoid that the flow of vibrational energy is artificially damped due to velocity rescaling by the thermostat.[132] However, NVE simulations are known to suffer from an spurious drift of the kinetic energy caused by the shadow Hamiltonian,[151] and therefore typically require a very short integration time step (say, $\delta t = 0.2$ fs). Following initial heating, moreover, the temperature of the total systems increases for a finite number of solvent molecules at long times, that is, the system does not return to the initial equilibrium state. As an alternative approach, it has been suggested[131] to perform NVT simulations where only the solvent is coupled to a thermostat, which uses a coupling time that is longer than typical energy transfer times.

We begin with the microcanonical ensemble, where the conservation of total energy E can be used as a simple measure of the accuracy of the simulation. Assuming equipartition of mean potential and mean kinetic energies, also the total kinetic energy is approximately constant. As a representative example, Fig. 2.2 shows the time evolution of the kinetic energies of PDZ3 protein and water solvent, $E_{\mathrm{P}}(t)$ [Eq. (2.2)] and $E_{\mathrm{S}}(t)$, respectively, adopting an initial temperature of $T_0 = 100$ K and a T-jump of $\Delta T = 600$ K. Following excitation at time $t = 0$, the protein energy is seen to decay on a 10 ps timescale, and the solvent energy rises accordingly. For times $t \gtrsim 20$ ps, however, we find that both protein and solvent energies increase. This clear violation of energy conservation occurs despite a quite small integration time step of $\delta t = 0.2$ fs.

To allow for a larger time step, we perform NVT simulations where only the solvent is coupled to a thermostat at temperature T_0. Figure 2.2 shows that for $t \lesssim 20$ ps we obtain the same decay of $E_{\mathrm{P}}(t)$ and corresponding rise of $E_{\mathrm{S}}(t)$ as in NVE. However, the NVT results behave reasonable also at longer times, as they are found to decay towards their equilibrium value which, though, is not reached within 100 ps. Results of similar quality are also obtained for residue energies [Eq. (2.1)] under NVE and NVT conditions (Fig. S7). After studying the performance of the NVT simulations obtained for various integration time steps δt and thermostat coupling times τ_{T} (Fig. 2.2), we use in all nonequilibrium simulations reported below $\delta t = 0.7$ fs and $\tau_{\mathrm{T}} = 10$ ps.

2.3.2 Cooling

Following the heating of the Azu residue to temperature $T_0 + \Delta T$ (see Methods), the excess energy $k_{\mathrm{B}}\Delta T$ flows to the remaining residues of the protein and to the solvent, both of which have initial temperature T_0. This results in a cooling of the hot protein in the cold solvent water, which is reflected in the decay of protein energy $E_{\mathrm{P}}(t)$ [Eq. (2.2)].

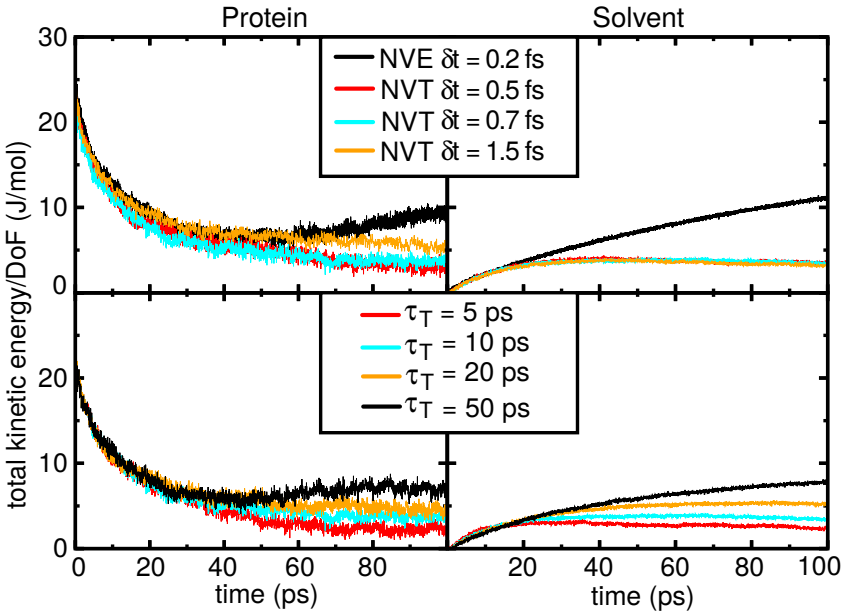


FIGURE 2.2: Choice of nonequilibrium simulation strategy. Following initial heating from $T_0 = 100$ K to $T_0 + \Delta T = 700$ K of the Azu residue, the time evolution of the mean kinetic energies of PDZ3 protein (left) and water solvent (right) is shown. While NVE simulations (black) exhibit a spurious energy drift for times $t \gtrsim 20$ ps, NVT simulations (colored) are found to decay to equilibrium energies. Moreover, NVT results for various integration time steps δt (top, using $\tau_T = 10$ ps) and thermostat coupling times τ_T (bottom, using $\delta t = 0.7$ fs) are compared.

Being a result of the complex interplay of Lennard-Jones and Coulomb interactions, the cooling rate particularly depends on the type of solvent and the force-field used.[131] While the T-jump excitation ($\Delta T = 600$ K) is determined by the $S_0 \rightarrow S_1$ excitation energy and therefore the same for all simulations, it is instructive to study various initial temperatures $T_0 = 10$, 100 and 300 K. We note that the initial energies of TrpZip2 and PDZ3 differ by some factor 10, since the same amount of excitation energy is distributed among 10 times more degrees of freedom in PDZ3. Interestingly, Fig. 2.3 reveals that the cooling time τ_C obtained from a single-exponential fit [Eq. (2.3)] depends significantly on temperature T_0 for both TrpZip2 and PDZ3. At 300 K, we find cooling times of ≈ 8 ps, which is in excellent agreement with reported experimental values of 6-9 ps.[44,82] At 100 and 10 K, on the other hand, τ_C is found to increase to ≈ 15 and 18 ps, respectively (cf. Table 2.1). Moreover, the mono-exponential decay apparently becomes questionable.

To explain these findings, we assume that dissipation of protein vibrational energy into the solvent occurs via rapid thermal collisions of the protein with water molecules. The translational motion of water can be illustrated by calculating the mean square displacement $\Delta r^2(t) = \langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \rangle$ of water, which shows diffusive behavior ($\Delta r^2(t) \propto Dt$) at 300 K (Fig. S8). Below the glass temperature of water at 120 K, the diffusion constant D is found to approach zero (Fig. S8), hence cooling can no longer be caused by the

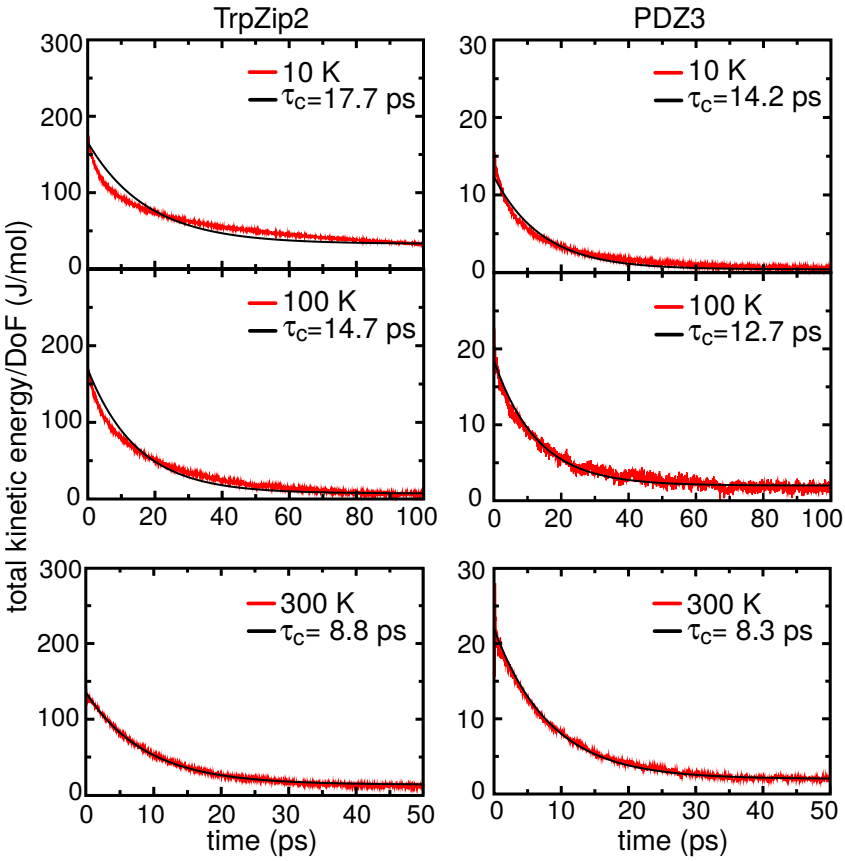


FIGURE 2.3: Cooling of the thermally excited protein in the cold solvent water as shown by the decay of protein energy $E_P(t)$ for TrpZip2 (left) and PDZ3 (right) at various initial temperatures. Black lines are single-exponential fits [Eq. (2.3)] of the data.

translational motion of water. However, cooling may be also achieved via librational vibrations of water, which appear to be the cause of the finite cooling times of 15 ps at 100 K. Interestingly, the cooling time remains approximately constant even at 10 K, which is in agreement with recent MD calculations on myoglobin, but in variance with accompanying experimental results.[152] We attribute this discrepancy to quantum-mechanical effects at low temperatures, which are not accounted for in our classical approach.[153,154]

We note in passing that we find only little effect of the solute on cooling, which is in line with previous studies.[131] Nonetheless, cooling times in TrpZip2 are generally somewhat longer than in PDZ3 (see Table 2.1). This difference may be attributed to the relatively large hydrophobic surface of TrpZip2 in comparison to PDZ3 (Fig. S9), which causes a weaker coupling between water and TrpZip2 and thus a slower energy transfer.

2.3.3 Energy transport in TrpZip2

We now turn to *intraprotein* transport of vibrational energy, and first consider the 12 amino-acid β-hairpin TrpZip2.[84] Starting at initial temperature $T_0 = 300$ K, we impulsively heat residue Azu3 by $\Delta T = 600$ K (see Methods), from where the energy propagates through the molecule via backbone transport and interresidue contacts. To obtain a sufficient signal-to-noise ratio, we used 5 000 trajectories and smoothed all time traces by a Gaussian filter (see the Supplementary Information for details and corresponding raw data in S1 - S4). Figure 2.4 displays the resulting time evolution of the residue kinetic energies [Eq. (2.1)]. Except for the energy of the initially excited heater residue Azu3 that decays rapidly, the residue energies are seen to rise on a picosecond timescale to a peak value before they decay towards zero within the simulation time of 50 ps. If the energy transport occurred predominantly via the peptide's backbone, we would expect a shift of the peak time with increasing sequence distance to the heater. This is the case for the adjacent residues on each side of the heater, Trp2 and Ser1 as well as Trp4, Glu5 and Asn6, which reach their peak energy within 3 - 5 ps.

Beginning with Gly7 at the turn of the β-hairpin, however, the situation becomes more involved, because effects of the energy transport via interresidue contacts become apparent. As indicated in Fig. 2.1, there are various types of contacts between residues pairs (3,10), (3,12), (1,12) and (5,8). Connecting both sides of the β-hairpin, these contacts provide a short-cut for the energy flow, which competes with the sequential backbone transport. For example, the energy of Trp9 reaches its peak before the Lys8 energy and exhibits a larger amplitude. This can be explained by contact transport between residues Azu3 and Aha10 and subsequent backbone transport to Trp9. The effects of contact transport are most prominent for the last three residues of the hairpin. Due to contacts (3,10) and (3,12), we find the earliest and strongest energy signal of Lys12, followed by Aha10 and then by Trp11.

It is instructive to study the structure of the main contacts of TrpZip2, see Fig. 2.5. For one, residues Azu3 and Aha10 are connected by two stable inter-backbone hydrogen bonds that stabilize the structure of the β-hairpin. Moreover, we find that the negative partial charge of N(1) of the Aha10 azido group interacts with the positively charged edges of Azu3, resulting in a direct side-chain contact of Aha10 with the heater. To estimate the effects of these two transport channels, Fig. 2.5 compares the time-dependent kinetic energy of the two NH·CO groups participating in the hydrogen bonds as well as the kinetic energy of the N_3 group of Aha10. From their similar appearance we conclude that the energy content of these two pathways is roughly similar. On the other hand, residue Lys12 is involved in two contacts, that is a stable hydrogen bond with Ser1 as well as a strong polar contact between its ammonium moiety and the negatively charged center of the Azu3 ring. Comparing again the corresponding kinetic energies of the connection atoms, we find that the side-chain contact to the heater residue results in larger energy transfer than the inter-backbone hydrogen bonds. As the heater

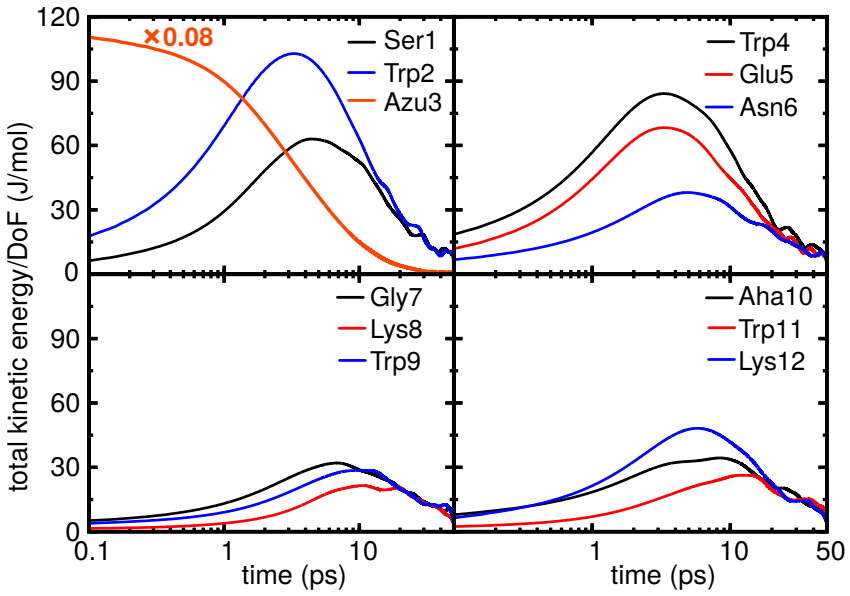


FIGURE 2.4: Time evolution of residue energies [Eq. (2.1)] in TrpZip2. Following initial T-jump excitation of residue Azu3, the energies of the adjacent residues, Trp2 and Ser1 as well as Trp4, Glu5 and Asn6 rise due to energy flow along the backbone. Residues at the other side of the β-hairpin (Gly7 - Lys12) receive energy via a complex interplay of backbone and contact transport.

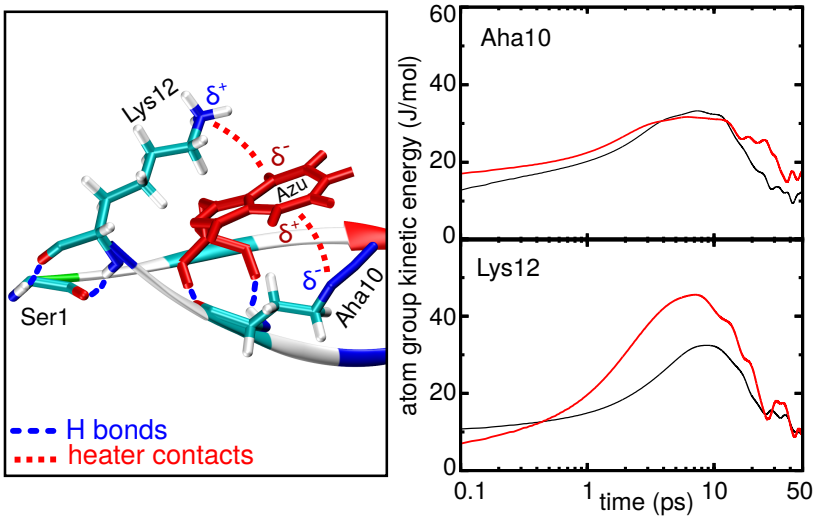


FIGURE 2.5: (Left) Structural characterization of various contacts bridging the β-strands of TrpZip2. (Right) Time-dependent kinetic energy of atom groups participating in the contacts of Aha10 and Lys12. Shown are energies of NH and CO atoms of the backbone hydrogen bonds (black) as well as energies of N_3 and NH_3 groups forming side-chain heater contacts of Aha10 and Lys12, respectively.

carries by far the most energy, and since the charged ring structure of azulene provides ample possibilities for polar contacts with other residues, which are stronger than the

conventional, ubiquitous van der Waals contacts, these side-chain "heater contacts" are generally found to be quite effective in spreading energy.

2.3.4 Energy transport in PDZ3

We are now in a position to discuss the energy transport in the allosteric protein PDZ3, which appears to be considerably more complex than in the small peptide studied above. For one, this is because the heater residue Azu is attached to the peptide ligand rather than directly to the protein. This means that the photoinduced excess energy first needs to transfer via non-covalent contacts from the ligand to the adjacent α_2 and β_2 regions of the protein, see Fig. 2.6. Moreover, PDZ3 exhibits a conformational transition in the binding pocket region (see Fig. 2.7), which may significantly affect the energy transfer of the system.

To obtain an overview of the energy flow in PDZ3, Fig. 2.6 shows the state-averaged residue energies which reveal that the energy first flows through the ligand, before it reaches the α_2-helix and the β_2-sheet, as well as a few more distant residues. Since the Azu heater is attached to the N-terminus of the ligand, the energy propagates through the ligand backbone and successively reaches all its residues within a few picoseconds. On the β_2 side, residues 325 - 329 directly face the ligand and form hydrogen bonds at positions 325 - 327 and 329, while Ile328 may form a side-chain contact directly with the Azu heater. Accordingly, residues 327 - 329 closest to the heater are seen to obtain more energy faster than residues 325 and 326. On the α_2 side, the ligand may form hydrogen bonds with His372 and Lys380. Moreover, heater contacts with His372, Glu373 and Ala376 exist. As a consequence, the latter three residue rapidly obtain a significant amount of vibrational energy. There are also a few contacts to neighboring residues that are more remote in sequence space. In particular, Glu331 and Phe400 receive energy via heater contacts, as well as Ile336 and Ser340. Apart from α_2 and β_2 adjacent to the ligand, energy is thus also found to transfer to β_3 (Ile336, Ser340), β_2-β_3 loop (Glu331, Glu334), and α_3(Phe 400).

As mentioned above, the energy flow in PDZ3 is complicated by the fact that the system exhibits a conformational transition in the binding pocket region, see Fig. 2.7. That is, due to the large flexibility of the ligand, the Azu residue may either turn to the β_2-sheet or the α_2-helix, which we refer to as state β and state α state, respectively (see Fig. S10 for a structural characterization of the two states). In state α, the heater preferably forms contacts with residues His372, Glu373 and Ala376. As a consequence, the energies of these residues rise significantly faster and higher in state α than in state β (Fig. 2.7). On the other hand, in state β the heater preferably forms contacts with residues Ile328, Glu331 and Phe400, whose energies rise significantly faster and higher in β than in α. While both states occur with similar probability in our simulations, recent NMR studies[155] showed that Azu forms contacts with residues 329 and 330, which indicates a preference for state β.

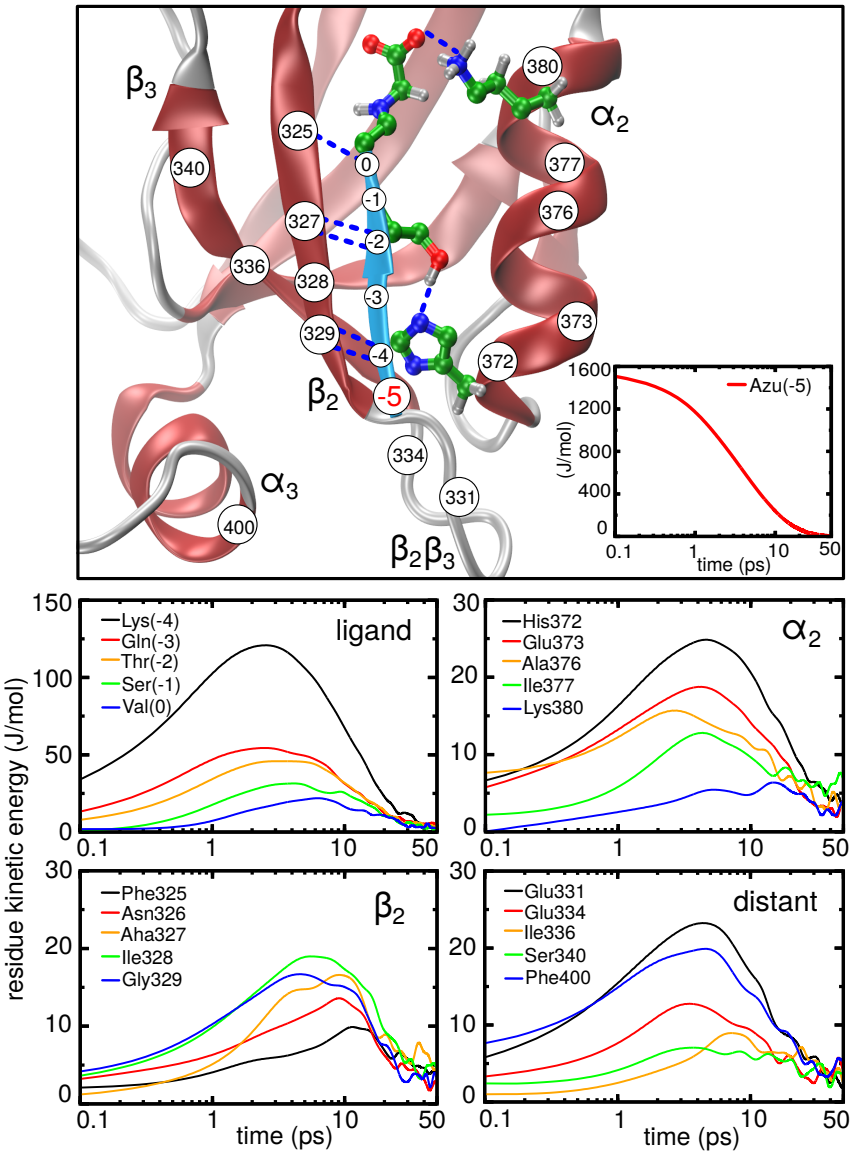


FIGURE 2.6: Energy transfer in PDZ3. (Top) Close-up view of the binding pocket, including the ligand (green) and the adjacent α_2 and β_2 regions of the protein (red). By convention, the ligand is labelled from -6 to 0, while the protein is labelled from 300 to 415. (Bottom) Time evolution of residue energies [Eq. (2.1)], reflecting the energy flow from the ligand via α_2 helix and β_2-sheet to more remote regions.

It is interesting to compare our findings to previous studies of the energy transport in PDZ3 domains.[128–130] For example, Wang et al.[130] recently employed nonequilibrium energy transport calculations using an elastic network model of PDZ3. Following initial excitation of residue His372, they reported energy flow to roughly the same regions of PDZ3 as reported above. Compared to an all-atom force field with explicit solvent description, a harmonic elastic network model using C_α distances clearly requires much less numerical effort. On the other hand, this highly idealized model naturally cannot describe issues such as different energy transport efficiencies obtained for various types

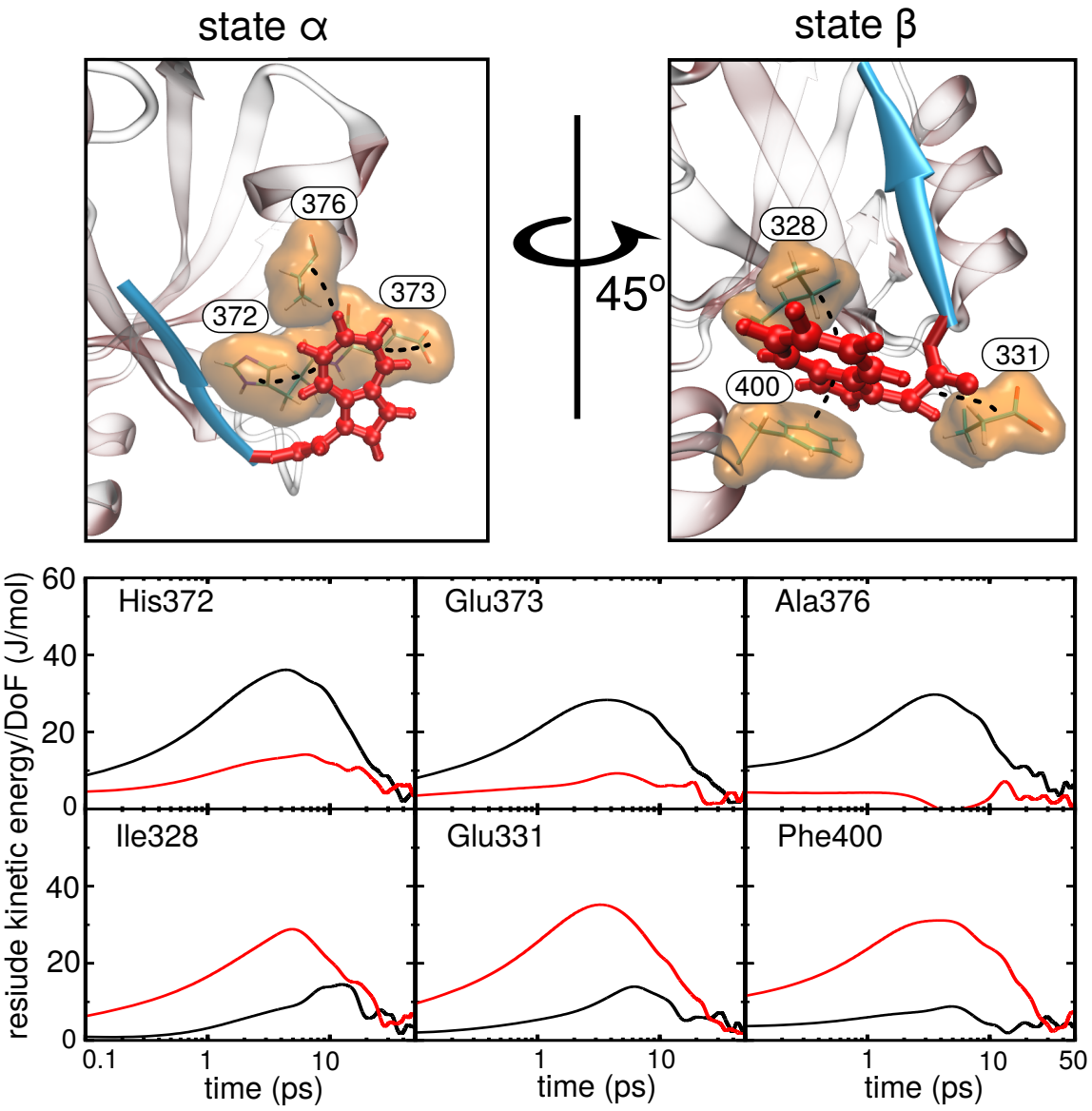


FIGURE 2.7: Effect of the $\alpha \leftrightarrow \beta$ conformational transition on the energy transport in PDZ3. (Top) Structure of the binding pocket region for conformational states α and β. (Bottom) Time-dependent energies of selected residues in state α (black) and state β (red).

of contacts, the effect of conformational transitions and the dissipation of energy due to cooling, all of which may significantly affect the energy transport measured in experiment. We also mention the detailed study of Ishikura et al.,[129] who performed all-atom equilibrium MD simulations to calculate the energy flow between any two residues of PDZ3. The resulting energy exchange network of PDZ3 indicates similar inter-residue contacts, e.g., between residues (330,334) in the β_2-β_3 loop, (371,374) of the α_2-helix, and (397,400) of the α_3-helix. However, this study is based on equilibrium MD simulations and without peptide ligand, and thus not directly comparable to our results.

2.4 Concluding remarks

Nonequilibrium MD simulations are arguably the most direct approach to compare theoretical predictions to time-resolved experiments on biomolecular energy transport. With this end in mind, we have developed an efficient simulation protocol that accurately mimics the excitation and probing steps of experiments such as performed by Bredenbeck and coworkers.[37,82] In particular, we have shown that NVT simulations (time step $\delta t = 0.7$ fs) with only the solvent coupled to the thermostat (coupling constant $\tau_T = 10$ ps) represents an efficient and accurate strategy that overcomes numerical problems associated with NVE calculations.[151] To achieve a sufficient signal-to-noise ratio of single-residue energies at 300 K, we have calculated up to 10 000 nonequilibrium trajectories. Employing this methodology, we are in a position to compare our numerical simulations to upcoming experiments on the energy transfer in TrpZip2 and PDZ3.[39] Moreover, our nonequilibrium simulations may serve as a reference for more approximate calculations of the energy flow.[86,129]

We have first studied the cooling of the photoexcited proteins in the cold solvent water. The nonequilibrium calculations were shown to quantitatively reproduce the experimental cooling times ($\approx$ 6-9 ps) at room temperature.[44,82] Below the glass temperature, the cooling was found to slow down by a factor two. This reflects the fact that the diffusional motion of water freezes and thus librational vibrations of water are the only remaining channel to dissipate energy into the solvent.

We have shown that in PDZ3 energy travels from the ligand via a complex network of inter-residue contacts and backbone transport to residues as remote than Phe400 located at α_3. Occurring on a picosecond timescale, the process is very fast and may be even faster when quantum effects are considered.[154] We have injected excess energy into the ligand and monitored its flow through the protein, indicating for pathways which may coincide with pathways of structural reorganisation. The latter are the basis of allosteric communication in PDZ3, which can be initiated by (un)binding of a ligand to PDZ3. Since allosteric communication in PDZ domain occurs on timescales of 10 ns to 10 µs,[14–16] it is obvious that energy transport and allostery are quite different phenomena. Nonetheless, conformational changes underlying allostery may be triggered by the same mechanical couplings (e.g., tertiary contacts) as used to transport energy.[31,33,75–78,124] Besides energy transport networks, though, numerous alternative suggestions of network models to predict allosteric pathways exist.[27,28,156] Due to different underlying assumptions, these models were found to yield quite different results for PDZ domains.[157] It remains to be shown to what extent energy transport pathways may be indicative of pathways of allosteric communication.

2.5 Supplementary Information

AZU Parameterization procedure

Force field parameters of Azulene (AZU) were obtained with Antechamber package.[142] For determination of atomic charges, the structures of AZU was optimized on B3LYP/6-31G* level using the GAUSSIAN g09 program suite.[143] Atomic charges were then computed as Mulliken charges from HF/6-31G* single point calculations. Point charges for MD calculations were then obtained from multiconformational restrained electrostatic potential (RESP)[158] calculations.

AZU Parameters

[AZU]

[atoms]

N	N	-0.41570	1
H	H	0.27190	2
CA	CT	-0.00240	4
HA	H1	0.09780	5
CB	CT	-0.03430	6
HB1	HC	0.02950	7
HB2	HC	0.02950	8
CG	CA	-0.16005	9
CD1	CA	-0.10650	10
HD1	HA	0.14050	11
CD2	CB	-0.04775	12
CE2	CB	-0.05925	13
CE3	CA	-0.02850	14
HE3	HA	0.14647	15
CZ2	CA	-0.03250	16
HZ2	HA	0.14022	17
CZ3	CA	-0.06000	18
HZ3	HA	0.13622	19
CH2	CA	-0.15625	20
HH2	HA	0.13872	21
C1	CA	-0.18600	22
H1	HA	0.14375	23
C2	CA	-0.15525	24
H3	HA	0.14047	25
C	C	0.59730	26
O	O	-0.56790	27

[bonds]

-C	N
N	H
N	CA
CA	HA
CA	CB
CA	C
CB	HB1
CB	HB2
CB	CG
CG	CD1
CG	CD2
CD1	HD1
CD1	C1
C1	H1
C1	CE2
CD2	CE2
CD2	CE3
CE3	HE3
CE3	C2
C2	H3
C2	CZ3
CZ3	HZ3
CE2	CZ2
CZ2	HZ2
CZ2	CH2
CH2	HH2
CH2	CZ3
C	O
C	+N

[angletypes]

CT	CA	CB	1	119.450	542.330 ; CB - CG - CD2 AZU GAFF	
CA	CB	CA	1	114.190	570.360 ; CG - CD2 - CE2 AZU GAFF	

[impropers]

-C	CA	N	H			
CB	CD1	CG	CD2			
CG	CE2	CD2	CE3			
CG	C1	CD1	HD1			
CD1	CE2	C1	H1			
CE3	CZ3	C2	H3			
CZ2	CD2	CE2	C1			
CZ2	CZ3	CH2	HH2			
CH2	CE2	CZ2	HZ2			
CH2	C2	CZ3	HZ3			
C2	CD2	CE3	HE3			
CA	+N	C	O			
N	CA	C	+N	105.4	0.75	1

Gaussian Smoothing procedure

For better representation of energy profiles, we employed a Gaussian filter for smoothing of raw data. We are mainly interested in overall rise and decay behaviour as well as the peak times. A logarithmic scale representation provides the clear observation of these observable. In order to clearly visualize the the energy profiles in logarithmic scale we implemented Gaussian filter with variable standard deviation σ. For a given system, first of all we define total number (N_σ) and a minimum (σ_{min}) and maximum (σ_{max}) value of σ. In all the cases we used 100 different values of σ, which vary according to time step: $\Delta\,\sigma\ =\ (\sigma_{max} - \sigma_{min})/N_\sigma$ starting from $\sigma_{min} = 1$ and ending at $\sigma_{max} = 120$. The procedure is applied for each given value of σ. In this way, we generated 100 smooth trajectories. From each smoothed trajectory we extract only a fragment of data; the length of that fragment follows a logarithmic progression and is directly related to the value of σ used for that particular trajectory. The final smoothed trajectory is then obtained through the concatenation of all these fragments.

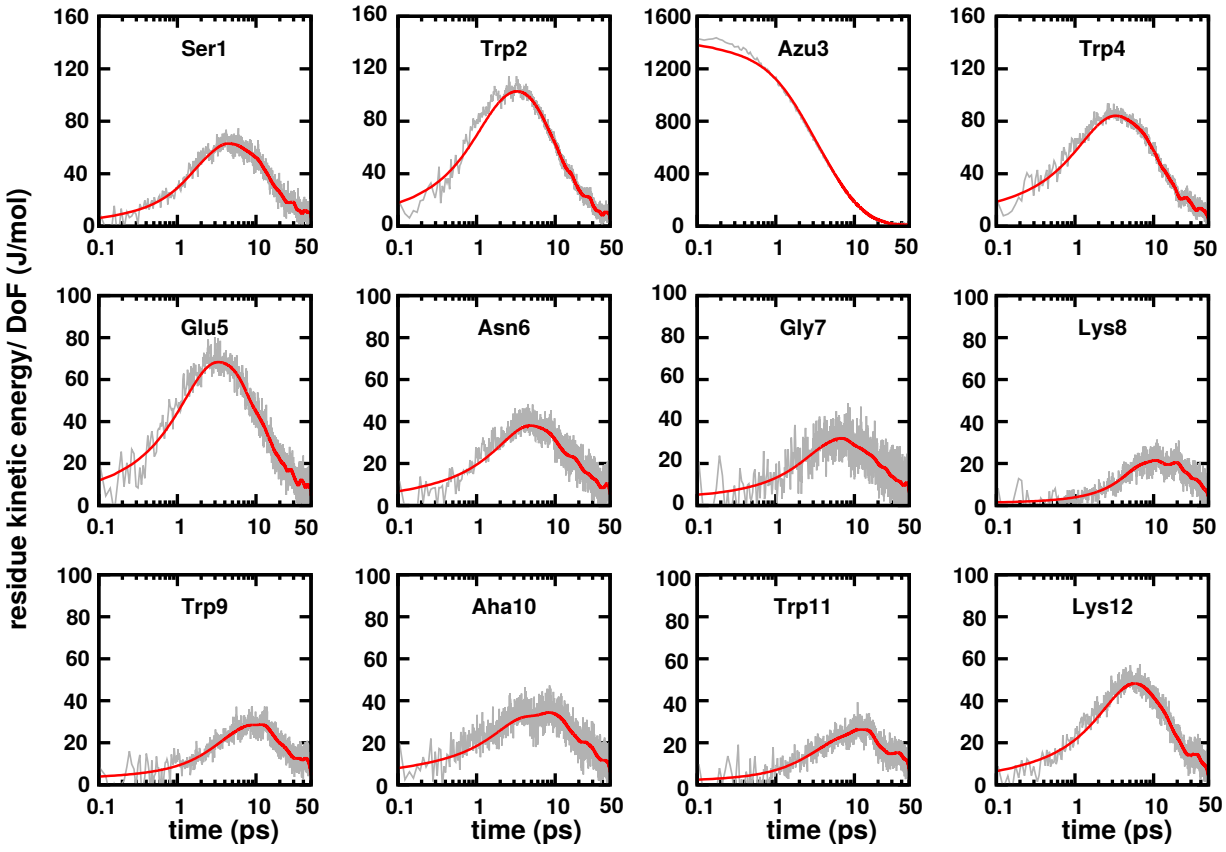


FIGURE S1: Time evolution of energy transport in TrpZip2. Filtered data (red) is shown on top of raw data (gray).

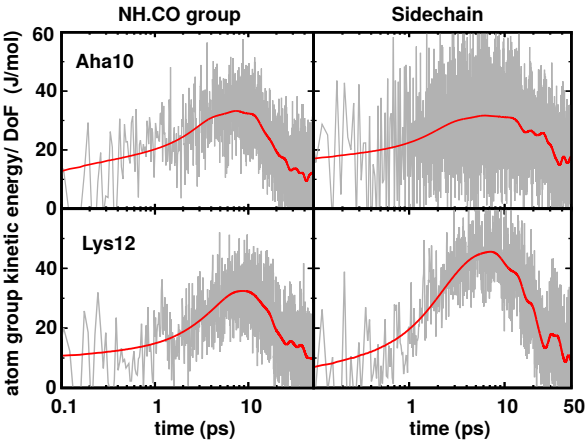


FIGURE S2: Comparison of energy transport through NH.CO group vs sidechain atoms involved in heater contacts of Aha10 and Lys12 are shown. Filtered data (red) is shown on top of raw data (gray).

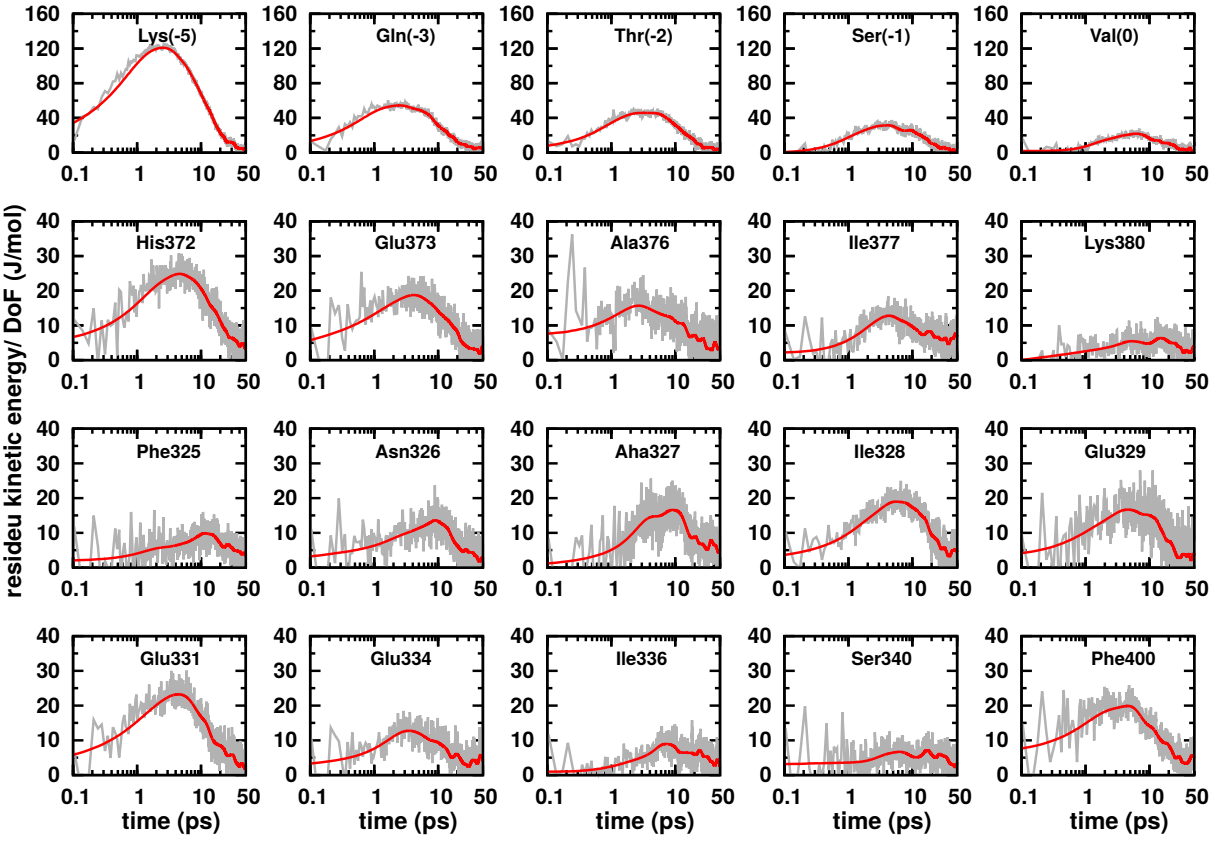


FIGURE S3: Time evolution of energy transport through residues of various regions of PDZ3. First row displays the residues of CRIPT peptide, second and third row display residues of α_2 and β_2 respectively. Last row shows the energy transport to far regions of PDZ3. Filtered data (red) is shown on top of raw data (gray).

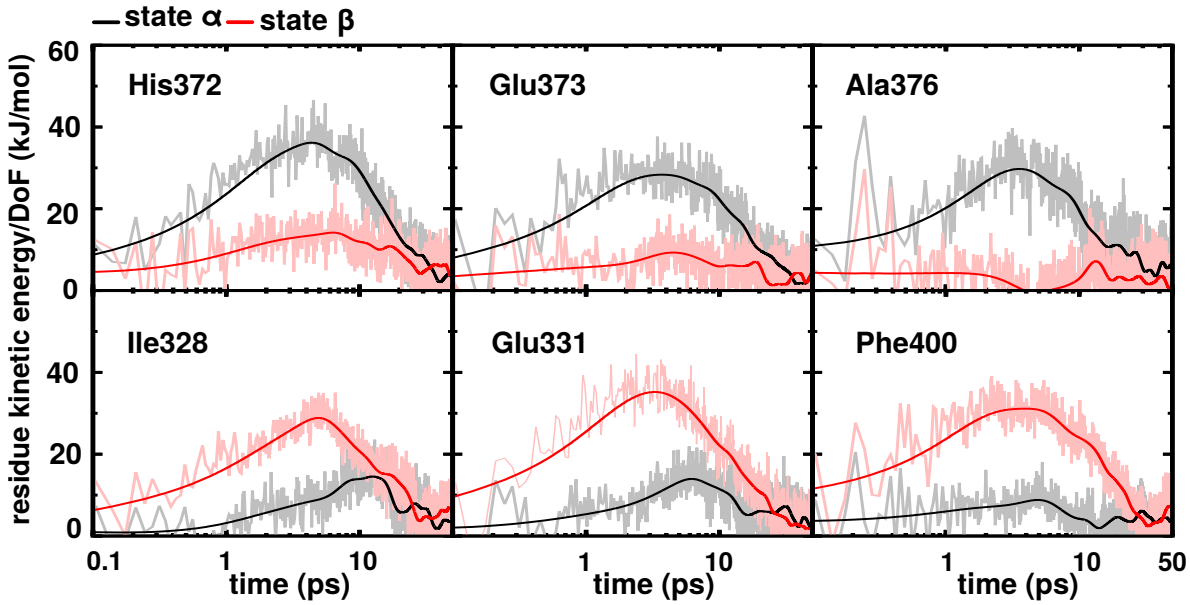


FIGURE S4: Comparison of energy transport through residues of α_2 and β_2 in state α at upper panels and in state β at lower panels.

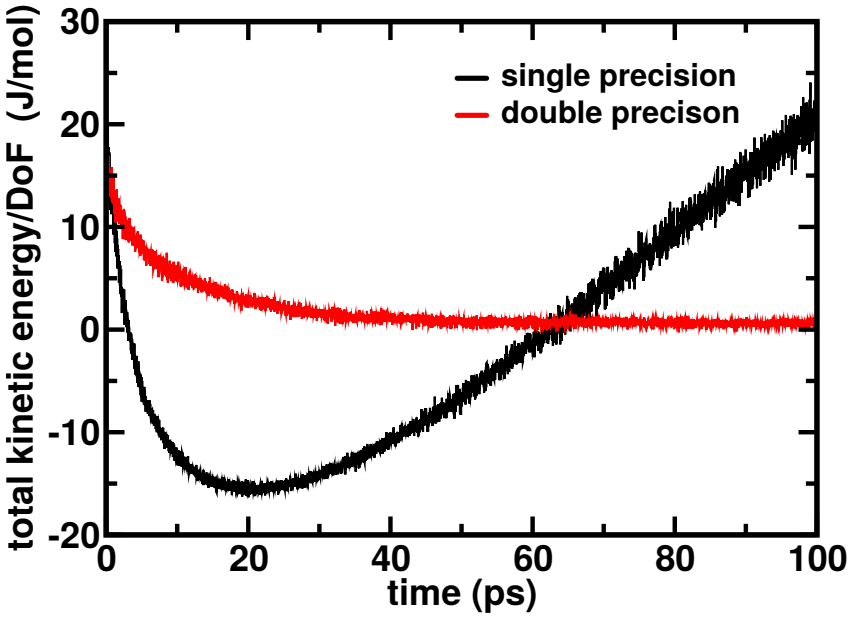


FIGURE S5: Time evolution of total kinetic energy of PDZ3 obtained by employing single precision (black) and double precision (red). Kinetic energies were calculated by averaging 150 trajectories in NVE ensembles. The time step of 0.2 fs was used for this comparison.

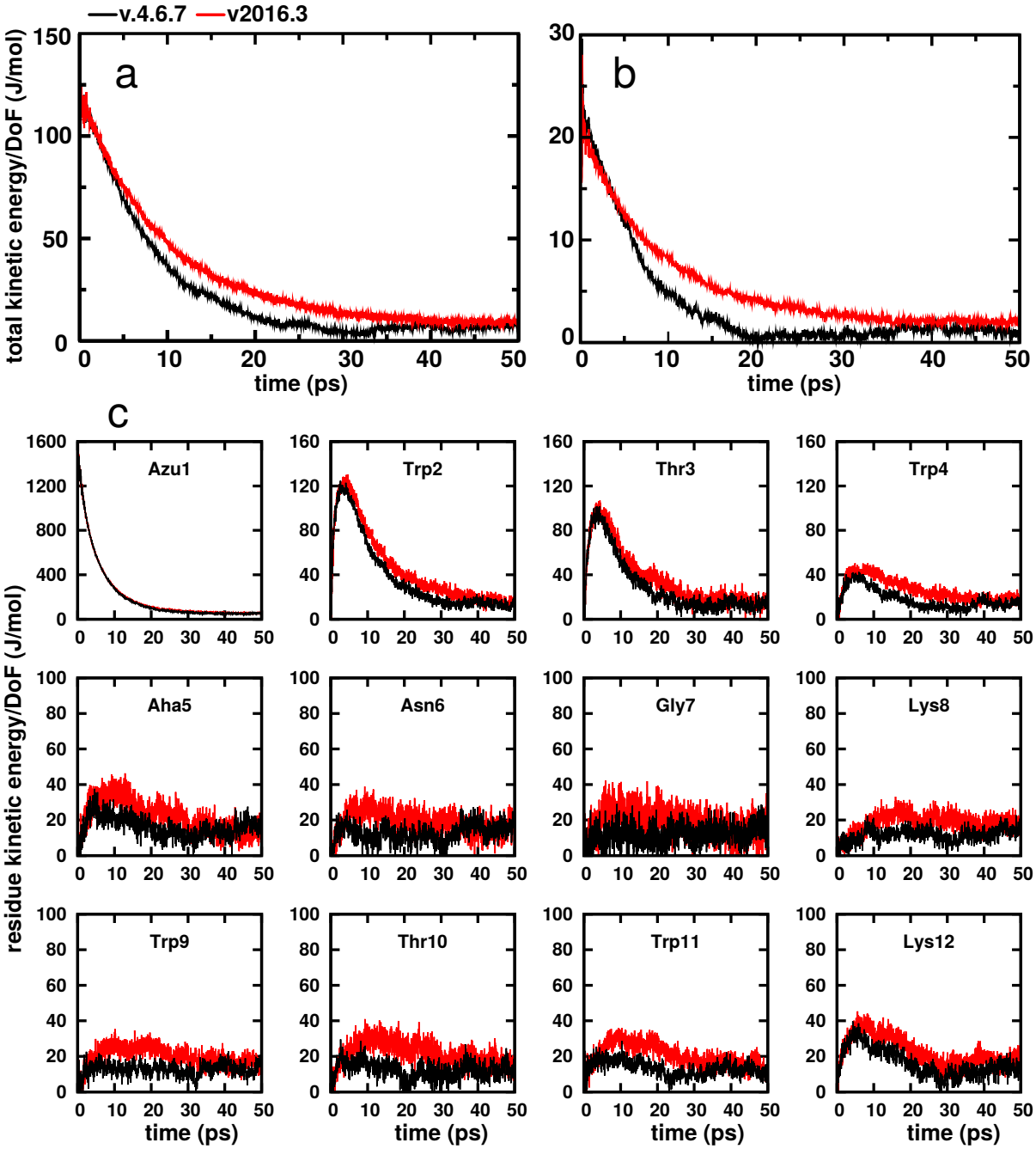


FIGURE S6: A comparison of kinetic energies obtained by employing v.4.6.7 and v2106.3 of gromacs. (a) total kinetic energy of TrpZip2,(b) total kinetic energy of PDZ3, (c) residue kinetic energy of TrpZip2. Same initial conditions were used for both cases.

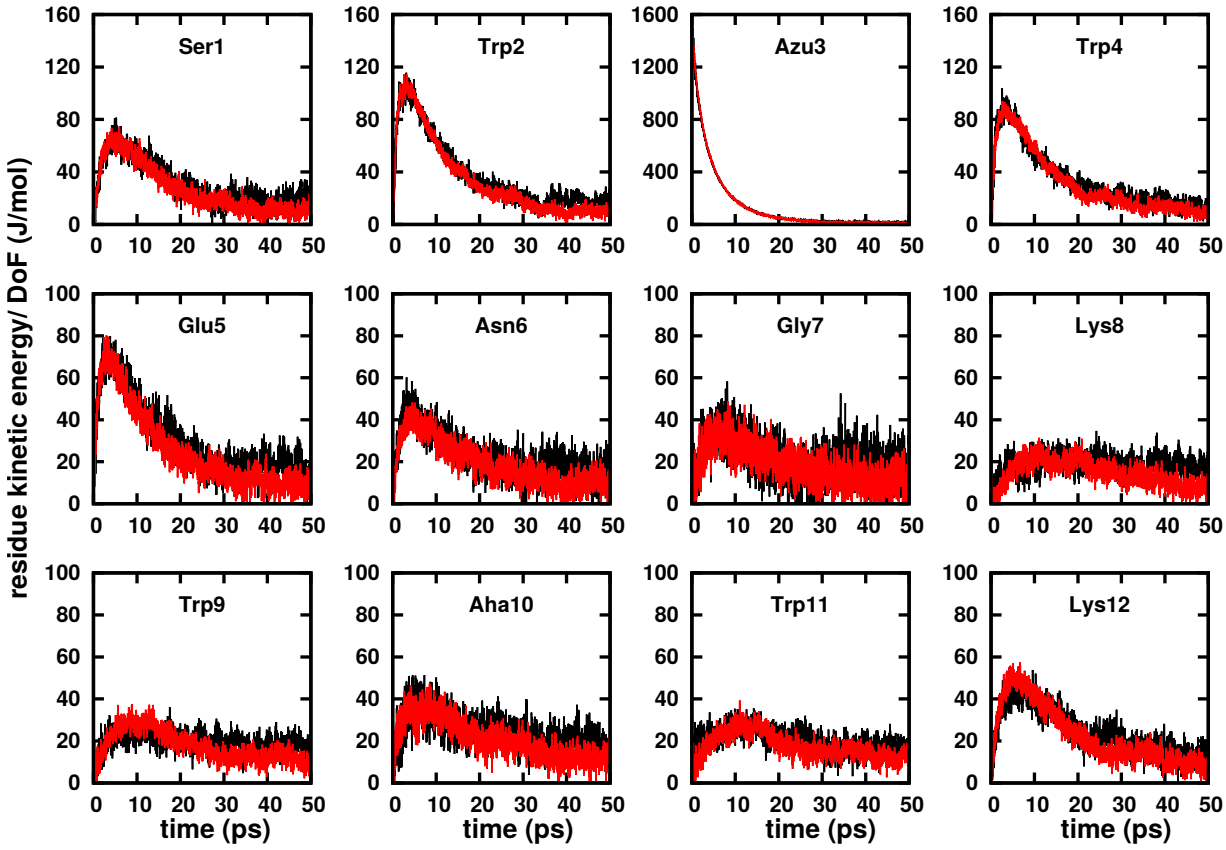


FIGURE S7: A comparison of residue kinetic energies of TrpZip2 in NVE (black) and NVT (red) ensembles. NVT ensemble at time step of 0.7 fs and solvent coupling constant of 10 ps reproduce the residue kinetic energies of NVE ensemble.

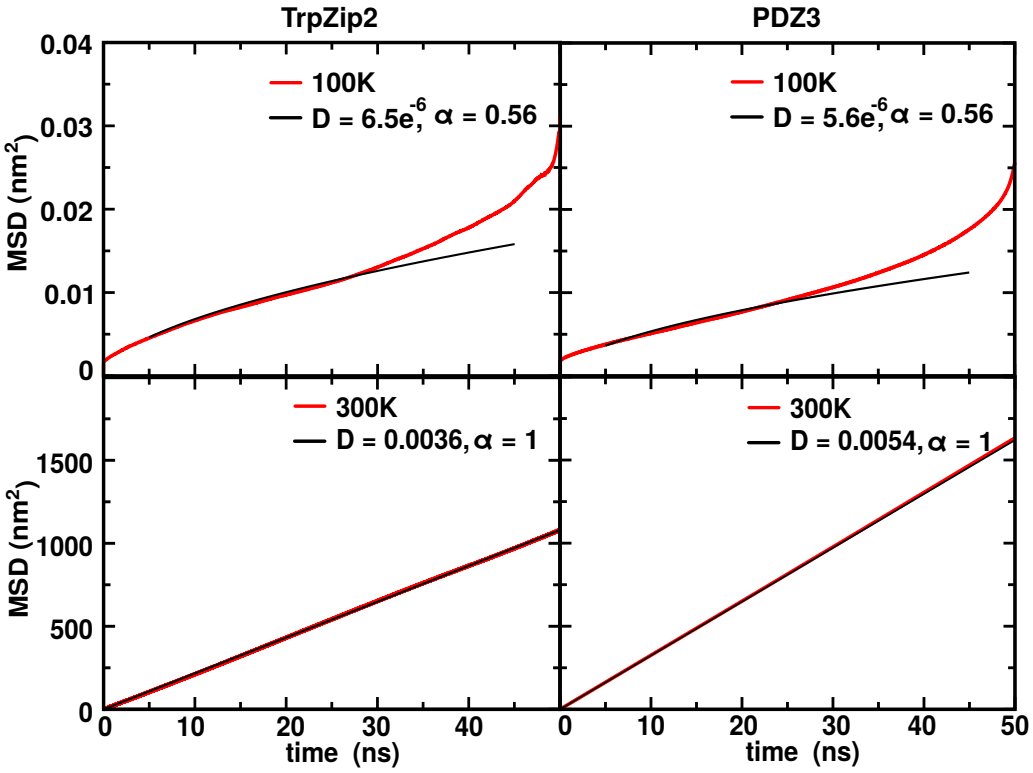


FIGURE S8: Mean square deviations of TrpZip2 and PDZ3 at 100K and 300K along with diffusion constants and α values describing the type of diffusion. (i.e $\alpha = 1$ equals normal diffusion, and $\alpha < 1$ is sub-diffusion).

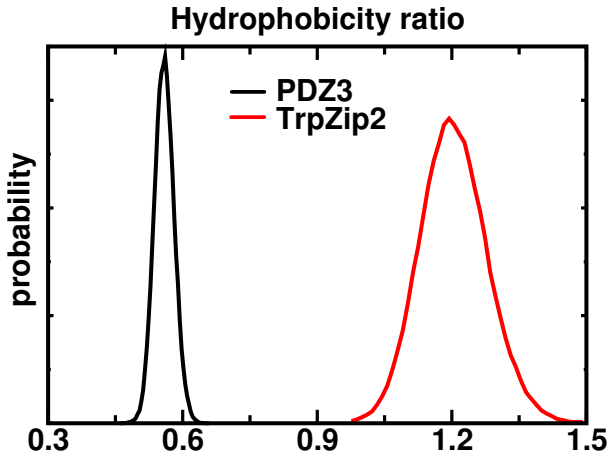


FIGURE S9: Hydrophobicity ratio (hydrophobic surface area divided by hydrophilic surface area) of PDZ3 and TrpZip2.

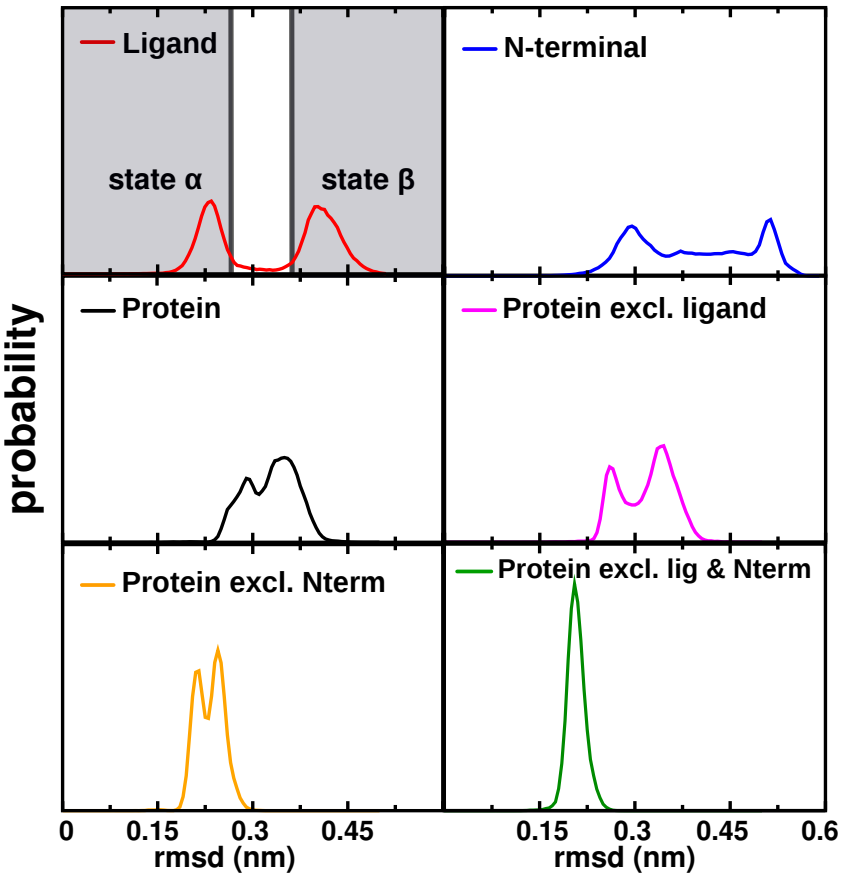


FIGURE S10: Probability distribution of root mean square deviation of various segments of PDZ3 domain. Ligand and N-terminus are identified as the most flexible parts of the complex. The ligand and N-terminus exists in two conformations. Ligand being equally distributed in two states namely state α and state β. The full complex including both N-terminus and ligand also exists in two conformations. While exclusion of the ligand and N-terminus from the calculations has a minor effect on the overall RMSD of complex the exclusion of both ligand and N-terminus from the calculations reveals that the RMSD of PDZ3 now exists only in one state. The RMSD analysis of the system therefore demonstrates that a global confromational change of PDZ3 is mainly due to local conformational changes of the ligand and the protein N-terminus

Master equation model to predict energy transport pathways in proteins
Here, I provided all the simulation data.

2.6 Master equation model to predict energy transport pathways in proteins

Progress in time-resolved vibrational spectroscopy has made it possible to monitor the flow of biomolecular energy in space and time.[41,74,123–125] Usually the energy is injected into the molecule via impulsive photoexcitation of some chromophore, e.g., a natural heme group[74] or a molecular photoswitch such as azobenzene.[41] Following ultrafast internal conversion into the electronic ground state, the vibrational energy may propagate along the protein backbone and via tertiary contacts such as hydrogen bonds, salt bridges and polar contacts. To measure the transient energy content of a particular protein residue, e.g., local C=O vibrations[41] or the azido group of unnatural amino acids can be employed.[159] In this way, recent experiments have been able to observe the anisotropic flow of vibrational energy in heme proteins[80] and PDZ3 domain.[82] Protein energy transfer is believed to indicate the long-range propagation of conformational change, which provides the basis for allosteric communication.[31–33,76,77]

Accompanying experimental studies, biomolecular energy flow has been described by atomistic molecular dynamics (MD) simulations.[31–33,38,76,77,130–134] Moreover, various network models of energy transport have been proposed,[85,86,128,129,135–138] which typically aim to predict the energy flow between specific parts (usually residues) of a protein. In particular, Buchenberg et al.[85,86] suggested a master equation

$$\frac{dE_j(t)}{dt} = \sum_i \left[k_{ij}E_i(t) - k_{ji}E_j(t)\right], \tag{2.4}$$

where E_i denotes the kinetic energy of residue i and k_{ij} represents the rate of energy transport from residue i to residue j. The model is valid in the case of diffusive energy transport,[86] which is typically found for solvated biomolecules.[41] The energy transport rates of the master equation can be obtained, e.g., from normal mode theory,[85] equilibrium MD simulations,[31] or nonequilibrium MD simulations.[38] If available, also experimental data or quantum energy transfer calculations[153,154] can be employed for that purpose.

Due to the large number of transport rates ($\propto N^2$ with N being the number of protein residues), a direct fit of the master equation to MD results is ill-defined and likely to yield nonphysical results. On this account, Buchenberg et al.[86] derived scaling rules, which aim to describe energy transport rates in terms of a few parameters. In particular, by exploiting the equivalence of the master equation and a discrete diffusion equation,

a scaling rule for the energy transport between two adjacent backbone residues i and $j = i \pm 1$ was derived,

$$k_{ij} = \frac{D_{\mathrm{B}}}{\langle x_{ij}^2 \rangle} \sqrt{\frac{f_j}{f_i}}, \tag{2.5}$$

where D_{B} denotes the backbone diffusion coefficient and f_i denotes the degrees of freedom of residue i. $\langle x_{ij}^2 \rangle$ represents the average square distance between every pair of atoms of residues i and j along covalent bonds, which reflects the average distance energy has to travel among every atom of both residues.[160] The factor $\sqrt{f_j / f_i}$ assures that the rates obey the detailed balance relation

$$k_{ij} f_i = k_{ji} f_j. \tag{2.6}$$

Moreover, a scaling rule for the energy transport via polar contacts was derived from a simple harmonic ansatz. To parameterize the scaling rules, Buchenberg et al.[86] performed nonequilibrium MD simulations of the small α-helical protein HP36. While the resulting master equation successfully reproduced the all-atom energy transfer simulations for HP36, the applicability of the model to other proteins, e.g., including β-sheets and different types of contacts, is not well understood. Moreover, to achieve sufficient signal-to-noise ratio, simulations in Ref.[86] were conducted at low temperatures (10 K), which raises questions on the generality of theory and resulting model parameters .

To extend the master equation model to the description of energy transport in general proteins at room temperature, in this paper we adopt recently performed nonequilibrium MD simulations of the anisotropic energy flow in proteins TrpZip2 and PDZ3 domain.[83] The study was inspired by experiments by Bredenbeck and coworkers, who used unnatural amino acids β-(1-azulenyl)-alanine (Azu) and azidohomoalanine (Aha) to site-specifically inject and probe vibrational energy in several peptides and proteins.[37,82] Apart from standard hydrogen bonds considered before, the modeling of these systems also requires to study the energy transport properties of aromatic contacts (e.g., formed by the Trp residues in TrpZip2), and of contacts due to cation-π and dipole-dipole interactions formed by the Azu heater. To aid the parameterization of the model, we furthermore performed additional energy transport simulations for various model systems, including Ala12, a simplified β-hairpin model AlaZip, and HP36, see Fig. 2.8.

Secondly, we want to exploit a key virtue of master equation (2.4), that is, the rate matrix $\{k_{ij}\}$ completely determines the time evolution of the system. Hence, also details of the dynamics, such as the most important energy transfer pathways between two points of the protein can be rigorously calculated from the rate matrix. Pathway calculations have been used in Markov state models to describe, e.g., the pathways of protein folding.[161] Here, we perform extensive Monte Carlo Markov chain simulations to explore the energy transport pathways in TrpZip2 and PDZ3.

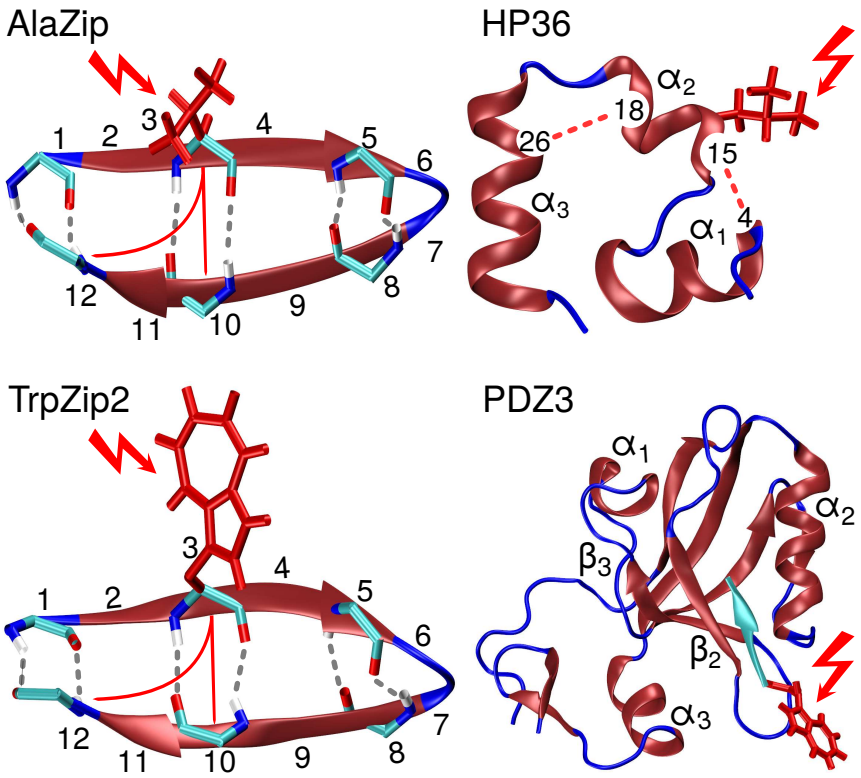


FIGURE 2.8: Scheme of the energy transport in the considered molecular systems, indicating the position of the initially excited heater residue (red), which at time $t = 0$ injects vibrational energy that is transported throughout the system via the protein backbone and interresidue contacts. For β-hairpins AlaZip and TrpZip2, direct side-chain contacts with the heater are indicated by red lines, and β-structure stabilizing hydrogen bond are indicated by gray lines. In the case of PDZ3, the Azu residue heats the ligand (cyan), from which the energy flows via various contacts to the protein.

2.7 Theory and Methods

2.7.1 MD simulations

Gulzar et al.[83] recently presented extensive nonequilibrium MD simulations of the energy flow in proteins TrpZip2[84] (PDB entry 1LE1) and PDZ3[121] (PDB entry 1BE9). All MD simulations were performed using GROMACS package v2016.3 (Ref.[48]), Amber99sb*ILDN forcefield,[139–141] and TIP3P water.[148] Following suitable equilibrium runs at $T_0 = 300$ K (100 ns length for TrpZip2, 5×1 µs for PDZ3), $N_{\text{traj}} = 5000$ and 10000 statistically independent initial structures were stored for the subsequent nonequilibrium runs of TrpZip2 and PDZ3, respectively. To mimic the initial heating of azulene via electronic excitation and subsequent ultrafast ($\sim$ 1 ps) internal conversion,[149,150] the resulting vibrational excitation was approximated by an instantaneous temperature jump, where the excess energy $k_B \Delta T$ is chosen to match the $S_0 \rightarrow S_1$ excitation energy of $\approx$ 2 eV, resulting in $\Delta T \approx 600$ K. Following the heating of Azu to $T_0 + \Delta T$, nonequilibrium MD simulations of 50 - 100 ps length were performed. It was found that NVT simulations

(time step $\delta t = 0.7$ fs) with only the solvent coupled to the thermostat (coupling constant $\tau_T = 10$ ps) represents an efficient and accurate strategy.[83]

To monitor the flow of vibrational energy from the heater residue through the protein, we consider the time evolution of the kinetic energy of the ith residue, $E_i^{\mathrm{kin}}(t) = \sum_j E_{i,j}^{\mathrm{kin}}(t)$, where the sum runs over all atoms j of residue i. The time-dependent expectation value of the kinetic energy per degree of freedom, $E_i(t)$, is calculated via an ensemble average over N_{traj} nonequilibrium trajectories,

$$E_i(t) = \frac{1}{f_i N_{\mathrm{traj}}} \sum_{n=1}^{N_{\mathrm{traj}}} E_i^{\mathrm{kin}}(n,t) - E_i^{\mathrm{eq}}, \tag{2.7}$$

where f_i denotes the degrees of freedom of residue i. Since $E_i^{\mathrm{eq}} = k_B T/2$ is the equilibrium energy per degree of freedom, $E_i(t)$ is expected to decay to zero at long times. The total kinetic energy per degree of freedom of a protein with M residues is then given by

$$E_P(t) = \frac{1}{M} \sum_{i=1}^{M} E_i(t). \tag{2.8}$$

For easier representation, all residue energies shown below were smoothed by a Gaussian filter, which used an adaptive width to account for the logarithmic representation of time.

To establish a general master equation model of biomolecular energy transport, it is important to base the parameterization of the model on several protein systems with different properties. Apart from TrpZip2 and PDZ3, we therefore used the above simulation protocol to also study HP36, Ala12 and AlaZip. Here, the α-helical protein HP36 (PDB ID 1UNC) was simulated at T_0 =300 K, in order to compare to previous results at low temperatures (10 K).[86] "Ala12" is build of 11 alanines and a valine at position 3, which was employed as heater residue. Undergoing frequent transitions between extended and helical structures (see Fig. S1), the system does not form contacts that are relevant for energy transport and therefore allows us to study backbone transport only. Using the same residues, "AlaZip" represents a model β-hairpin that was equilibrated at 100 K in a hairpin structure. Following heating of the peptide to 300 K within 3 ps and subsequent equilibration for 12 ps, the peptide stayed in β-hairpin structure during the subsequent 50 ps nonequilibrium MD runs. Lacking the aromatic contacts formed by the Trp residues in TrpZip2 as well as the heater contacts of Azu, AlaZip is used to focus on the energy transport along the hydrogen bonds connecting the β-sheet. In all cases, $N_{\mathrm{traj}} = 5\,000$ nonequilibrium energy transport simulations were performed. As an overview, Table 2.2 comprises all simulated protein systems together with the type of heater residue.

TABLE 2.2: List of considered molecular systems, type of heater residues, and resulting master equation parameters including inverse heating rate $1/k_{\mathrm{h}}$, solvent cooling time $1/k_{\mathrm{ps}}$, solvent back-transfer time $1/k_{\mathrm{sp}}$, and contact times $\tau_{i,j}=1/k_{i,j}$ (all in units of ps).

System	Ala12	AlaZip	TrpZip2	HP36	PDZ3
Heater	Val3	Val3	Azu3	Leu16	Azu(-5)
$1/k_{\mathrm{h}}$	1.7	1.7	5.9	3.1	5.9
$1/k_{\mathrm{ps}}$	7.9	8.7	8.3	8.8	10
$1/k_{\mathrm{sp}}$	210	250	120	120	60
contact times	-	$\tau_{0,10}=21$ $\tau_{3,10}=10$ $\tau_{0,12}=31$ $\tau_{1,12}=17$ $\tau_{5,8}=13$	$\tau_{0,10}=125$ $\tau_{3,10}=5.9$ $\tau_{0,12}=43$ $\tau_{1,12}=7.7$ $\tau_{5,8}=9.1$ $\tau_{0,5}=71$	$\tau_{4,15}=9.1$ $\tau_{18,26}=17$	$\tau_{0,328}=50$ $\tau_{0,331}=67$ $\tau_{0,334}=130$ $\tau_{0,400}=37$ $\tau_{-2,372}=20$ $\tau_{-4,331}=50$ $\tau_{-4,372}=23$ $\tau_{329,372}=33$

2.7.2 Master equation

As explained in the Introduction, a typical energy transport experiment consists of the impulsive excitation of some protein residue (henceforth referred to as "heater"), the propagation of the energy throughout the biomolecule via backbone and contact transport, and the probing of the local temperature[126] at specific residues. Moreover, the hot protein is cooled by the surrounding solvent molecules at temperature T_0, such that the system is at thermal equilibrium at long times. In master equation (2.4), all these processes are modeled by transport rates $k_{ij} \equiv k_{i\to j}$. Here, indices i and j run from 1 to N to describe energy transfer between the N residues of the protein. The initial excitation occurs via a heater unit with index $i=0$, the solvent degrees of freedom are collectively accounted for by the index $i=N+1$. In the following, we discuss the theoretical basis to determine these transport rates.

We begin with the modeling of the heating process, which in the nonequilibrium MD simulations is mediated by the initial excitation of the heater side-chain. In the master equation, the subsequent energy transfer from the heater side-chain to the backbone atoms of the heater is described by the heater rate k_{h}. Denoting the heater side-chain with index $i=0$ and the residue number of the heater in the considered system by index $i=n$ (e.g., in TrpZip2, $n=3$), we define rates $k_{0i} = k_{\mathrm{h}}\delta_{in}$ and corresponding back rates k_{n0} according to Eq. (2.6).

As a simple model of the cooling process, we assume that the energy of every protein residue dissipates into the solvent (index $i=N+1$) with rate k_{ps}, that is, we have $k_{i,N+1} = k_{\mathrm{ps}}$ for $i = 0,\ldots,N$. To reach thermal equilibrium at long times, we also need to invoke a small back-rate $k_{N+1,i} = k_{\mathrm{sp}}$. While more sophisticated models can be considered,[162] (e.g., k_{ps} could be a function of the solvent accessible surface of a particular residue), for the present applications a single solvent rate including back-rate turned

out to be sufficient.

The energy transport inside the protein occurs through its backbone and via interresidue contacts. Adopting the scaling rule in Eq. (2.5), in the former case we only need to determine the backbone diffusion coefficient D_B, since the number of degrees of freedom f_i of all residues is known and the average square distance $\langle x_{ij}^2 \rangle$ is readily obtained from the molecular structure. In principle, D_B may depend on the secondary structure (e.g., α-helical, turn or β-sheet) and on temperature.

2.7.3 Models of contact transport

Due to the many different types of interresidue contacts, a general description of contact transport rates turns out difficult. In the case of geometrically simple contacts, Buchenberg et al.[86] proposed a harmonic model that yields for the contact transport rate

$$k_{ij}^C = \frac{B_C}{\langle \delta q_{ij}^2 \rangle} \sqrt{\frac{f_j}{f_i}}, \tag{2.9}$$

where B_C is the ballistic contact transport constant and $\langle \delta q_{ij}^2 \rangle = \langle q_{ij}^2 \rangle - \langle q_{ij} \rangle^2$ represents the variance of the contact distance. This scaling rule was shown to work well for all polar contacts (including hydrogen bonds) in HP36.[86] Calculating $\langle \delta q_{ij}^2 \rangle$ from a short equilibrium MD simulation, only the single constant B_C was determined to model all contact rates.

As shown in Appendix A.1, though, the above model has several limitations. For one, the derivation assumes ballistic motion of two coupled oscillators. This is in variance with our assumption of diffusive backbone transport, although the difference between ballistic and diffusive motion should be small for contact transport involving only a single transport step. More importantly, the rate is found to depend on the way the contact atoms are connected to the rest of the protein. That is, rigidly connected contacts lead to smaller rates than loosely bound contacts. This may explain our findings below that the prominent intrastrand hydrogen bonds between β-sheets lead to relatively modest contact transport rates. To allow for a single transport constant B_C, Buchenberg et al.[86] tacitly assumed that these connections are similar for all considered contacts, which in general may not be valid.

Alternatively, we may describe both backbone and contact transport as diffusion processes. To this end, we adopt a multidimensional model that describes backbone transport and contact transport as diffusion processes in different directions using different energy diffusion constants. In direct analogy to scaling rule (2.5) for backbone transport, we then obtain for the contact transport rate

$$k_{ij}^C = \frac{D_C}{\langle q_{ij}^2 \rangle} \sqrt{\frac{f_j}{f_i}}, \tag{2.10}$$

where D_C denotes the contact diffusion constant and $\langle q_{ij}^2 \rangle$ represents the mean square distance of the two contact atoms [rather than the corresponding variance as in Eq. (2.9)]. In fact, Reid et al.[138] has recently found for myoglobin that the rates of various charged contacts rather obey Eq. (2.10) than Eq. (2.9). We note that the electrostatic force mediating a polar contact also scales with the inverse square distance, which reflects the electrostatic origin of the contact transport rate.

We finally mention several technical issues: (1) Besides geometrically simple contacts such as hydrogen bonds, also more complex cases such as aromatic contacts and contacts due to cation-π and dipole-dipole interactions exist. Here the interaction typically involves a number of atoms (e.g., of an aromatic ring) and can therefore not be appropriately described by a single distance. (2) In cases where two residues are coupled via two or more simultaneously existing contacts (e.g., a double hydrogen bond connecting a β-hairpin), the total transport rate is assumed to be given by the sum of the rates of the individual contacts. As a consequence, the scaling rules predict that the total rate scales with the sum of the inverse variance or mean square distance. (3) Some contacts may be present for only some percentage P of the time. Since the scaling rules apply only for interacting residues, we calculate variance and mean square distance only for the connected state and multiply the resulting rate with weighting factor P.

2.7.4 Parameterization

As discussed above, the parameterization of master equation (2.4) requires us to obtain the heater rate k_h, the solvent rates k_{ps} and k_{sp}, the backbone diffusion coefficient D_B, and all contact rates k_{ij}^C. In principle, we may find these parameters by minimizing the root mean squared deviation (RMSD) between the residue energies $E_j(t)$ of the master equation and the corresponding energies $E_i^{MD}(t)$ from the nonequilibrium MD data,

$$\mathrm{RMSD} \propto \left[\sum_{i,t} \left(\frac{E_i^{MD}(t) - E_i(t)}{E_i^{MD}(t)} \right)^2 \right]^{1/2} . \tag{2.11}$$

Here the sum runs over all residues $i = 0, \ldots N+1$ and over all simulation times $t = t_{min} + n\delta t$, where we choose $\delta t = 0.0175$ ps to include $\approx 10^4$ data points in the fit for each residue, and $t_{min} = 0.1$ ps to discard extreme energy fluctuations of the initial phase of the nonequilibrium MD simulations. Proceeding this way, Buchenberg et al.[86] employed a backtracking algorithm to obtain a global fit of the model parameters for HP36. In direct extension of this work, here we aim to determine energy transfer rates for a number of model proteins, including α and β secondary structures and a variety of interresidue contacts. As the success and quality of the resulting global fits may critically depend on the initial values used, it is helpful to first identify rough but physically reasonable values for all main transfer rates by trying to fit them individually.

In a first step, the solvent decay rate k_{ps} of each residue and the corresponding back rate k_{sp} are determined. Since the number of degrees of freedom of a protein residue is expected to be much smaller than the (unknown) effective number of degrees of freedom of the solvent coupling back to that residue, the detailed balance condition in Eq. (2.6) suggest that the back rate k_{sp} is at least an order of magnitude smaller that the forward rate k_{ps}. Hence k_{ps} is roughly given by the initial decay rate $1/\tau_S$ of the total energy of the protein in Eq. (2.8). Assuming for the moment a two-state model for protein and solvent and starting with $k_{ps} = 1/\tau_S$ and a rough value for the back rate (say, $k_{sp} = 0.05\,k_{ps}$), optimization of the two rates readily yields reasonable results.

To model the initial excitation of the protein and the subsequent energy transport via the backbone, we next want to determine the heater rate k_h (defined in Sec. 2.7.2) and the backbone diffusion coefficient D_B (defined in Eq. (2.5)), respectively. Assuming some approximate values of all other parameters, these quantities are optimized to reproduce the initial decay of the energies of the heater and its adjacent residues, as well as the overall propagation of the residue's peak energy along the backbone.

To describe contact transport, we need to identify the interresidue contacts of the systems that are relevant for energy transport. By calculating interresidue distances between the two closest respective atoms and using a suitable distance cut-off (e.g., ≤ 4.5 Å), we obtain a contact map of the system.[52] Note that we only need to consider contacts that are close enough to the heater to receive a significant amount of excitation energy. Moreover, it was found[86] that typically only polar contacts transport energy efficiently. That is, even nonpolar contacts arising from prominently stacked aromatic rings as in HP36 and TrpZip2 turn out to be negligible. The argument does not hold, though, for nonpolar contacts involving atoms close to the heated moiety. This is because the energy content of the heater is typically more than a factor 10 larger than for the other residues, and therefore also relatively weak contacts may become important.

To study the applicability of the above introduced scaling rules, we first fitted all contact rates and studied their correlation with Eqs. (2.9) and (2.10). The resulting scaling rule parameters B_C and D_C and the extent of their validity represent a central result of this work. A nontrivial complication arises, if two residues are connected by two adjacent energy transfer channels. For example, residues 5 and 8 of TrpZip2 are connected via the backbone as well as via intrastrand hydrogen bonds (see Fig. 2.1). Since the corresponding backbone transport rate is significantly higher than the contact rate $k_{5,8}$, a large variety of choices for the latter may give quite similar fits of the residue energies. A similar problem occurs for residues 3 and 10 of TrpZip2, which are connected via intrastrand hydrogen bonds as well as via a direct heater contact between the side-chains. To minimize the ambiguity associated with the fitting of the two contact rates, in these cases we first estimate the relative contribution of the two competing channels and use this information as additional input in the subsequent fit, see the Supplementary Material.

Given these initial estimates for all transfer rates, we perform for each molecular system a series of global fits to optimize all (or a subset) of parameters, using a backtracking algorithm as well as visual inspection. In particular, the generic parameters (i.e., backbone diffusion coefficient D_{B}, heater rate k_{h} for a specific type of heater residue, and contact scaling rules constants B_{C} and D_{C}) are adjusted such that we achieve a consistent model that is valid for all considered systems.

As a note of caution, we wish to mention that the RMSD landscape (2.11), considered as a function of the above described model parameters, is quite rugged and contains many local minima. Hence we typically find a range of model parameters resulting in similar quality of the fits. This is because of a number of uncertainties, including (1) the statistical convergence of the underlying nonequilibrium MD data, which are particularly noisy at far distances from the heater and at long times, (2) the fact that a master equation may only represent a rough approximation of the complex time traces obtained from MD, (3) the assumption of purely diffusive transport, while ballistic transport may become important for homogeneous systems such as Ala12 and at low temperatures, (4) the assumption of a common solvent decay rate for all residues, thus neglecting their possibly different solvent exposure, (5) the assumption of a common contact diffusion constant, disregarding the atomic details of various contacts, and (6) problems associated with competing contacts which introduce some ambiguity.

2.7.5 Identification of energy transport pathways

As explained in the Introduction, in a typical energy transport experiment we inject at time $t = 0$ vibrational energy into a specific residue i, and probe at a delayed time t the energy content $E_j(t)$ of some distant residue j. The energy flow from residue i to residue j may in general occur via several pathways, e.g., one pathway proceeding exclusively via the backbone and other pathways using a combination of backbone and contact transport. In practice, we are interested in the most efficient pathways that at a given time t have carried most of energy from i to j. Since the rate matrix $\{k_{ij}\}$ of master equation (2.4) completely accounts for the dynamics of the system, it also contains all information on energy transport pathways and their efficiency. Most straightforwardly, we obtain this information by running Markov chain Monte-Carlo simulations of the N-state system. At each step, a random number is drawn which determines if the system remains in the current state or changes to some other state. In this way, we sample a stochastic trajectory in state space according to rate matrix $\{k_{ij}\}$. By counting how often the system has propagated from i to j along each possible pathway, the weights of these pathways are readily calculated.

To obtained converged results, here we typically ran $6 \cdot 10^7$ Markov chain Monte-Carlo simulations of lengths from 0.1 to 50 ps. To reduce the number of pathways, we lumped together pathways that only differ by "loops." For example, pathways $1 \to 2 \to 3$ and $1 \to 2 \to 4 \to 2 \to 3$ would be combined into a single pathway $1 \to 2 \to 3$ with a

weight given by the sum of the two individual path weights. Moreover, we model the cooling effect of the solvent by discarding all pathways that include a transition into it.

2.8 Results and Discussion

To derive a general and consistent set of parameters that determine the rate matrix $\{k_{ji}\}$ of master equation (2.4), we start with the simple model system Ala12, that allows us to focus on backbone transport and energy dissipation into the solvent, as well as the α-helical protein HP36. Next we consider two types of β-hairpins, AlaZip and TrpZip2, which exhibit energy transport via various types of contacts. Employing the parameters obtained for these models, we are able to discuss the anisotropic energy flow in the allosteric protein PDZ3.

2.8.1 Backbone transport and cooling

The model peptide "Ala12" (11 alanines and a Val3 as heater residue) undergoes frequent transitions between extended and helical structures (Fig. S1), and therefore hardly forms contacts that are relevant for energy transport. Hence the energy flow in Ala12 is solely due to backbone transport and cooling of the peptide via the solvent. As an overview, Fig. 2.9 shows the time evolution of the residue energies $E_j(t)$ [Eq. (2.7)] obtained from the nonequilibrium MD simulations. Except for the energy of the initially excited heater residue Val3 that decays rapidly, the residue energies are seen to rise on a picosecond timescale to a peak value. As expected for backbone transport, we observe a shift of the peak time with increasing sequence distance to the heater. Due to the energy dissipation into the solvent, the residue energies decay toward zero within the simulation time of 50 ps.

In master equation (2.4), the backbone transport is accounted for by the heater rate k_h and the backbone diffusion coefficient D_B, while the energy dissipation into the solvent is described by the rates k_ps and k_sp. Adopting $1/k_\mathrm{h} = 1.7\,\mathrm{ps}$, $D_\mathrm{B} = 1.1\,\mathrm{nm}^2/\mathrm{ps}$, $1/k_\mathrm{ps} = 7.9\,\mathrm{ps}$ and $1/k_\mathrm{sp} = 210$ ps, Fig. 2.9 reveals that the master equation model reproduces the overall time evolution of the simulation results quite closely.

It is instructive to study the sensitivity of these results with respect to the model parameters. Concerning the backbone transport, we find that several combinations of heater rate k_h and the backbone diffusion coefficient D_B give similar results (e.g., $1/k_\mathrm{h} = 2.5\,\mathrm{ps}$ and $D_\mathrm{B} = 1.4\,\mathrm{nm}^2/\mathrm{ps}$, see Fig. S2). Here we choose $D_\mathrm{B} = 1.1\,\mathrm{nm}^2/\mathrm{ps}$ which best approximates the observed peak time of the residue energies, and $1/k_\mathrm{h} = 1.7\,\mathrm{ps}$ as corresponding best fit. Insertion of $D_\mathrm{B} = 1.1\,\mathrm{nm}^2/\mathrm{ps}$ in Eq. (2.5) yields an energy transport time between two adjacent alanine residues of ≈ 0.43 ps, which is in line with previous work.[38]

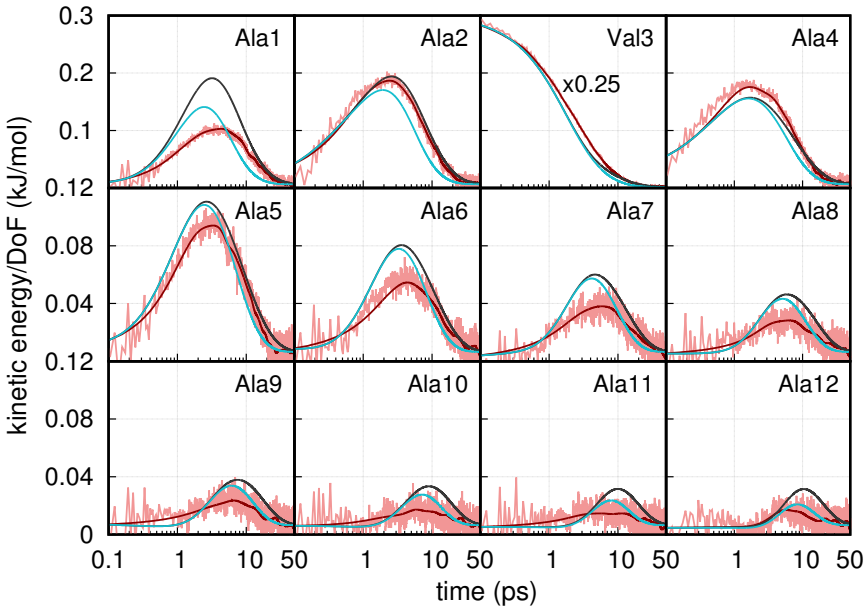


FIGURE 2.9: Time evolution of residue energies [Eq. (2.7)] of Ala12, obtained from nonequilibrium MD simulations (raw data in light red, Gaussian smoothed data in dark red). Following initial T-jump excitation of residue Val3, the energies of the adjacent residues rise due to energy flow along the backbone. Black lines represent results from the master equation model, blue lines correspond to an improved model using enhanced terminal solvent dissipation rates (see text).

Concerning the cooling of the protein in the solvent, we find that the solvent decay rate k_{ps} and the corresponding back rate k_{sp} can be readily obtained from a fit of the total protein energy, see Sec. 2.7.4. Using various values of k_{ps} and k_{sp}, Fig. S3 reveals that we get reasonable results for 7.5 ps $\lesssim 1/k_{\mathrm{ps}} \lesssim$ 8.5 ps and 150 ps $\lesssim 1/k_{\mathrm{sp}} \lesssim$ 300 ps, showing that particularly the value of the back rate is not very critical. Note the factor $k_{\mathrm{ps}}/k_{\mathrm{sp}} \approx 30$ reflects the different number of degrees of freedom of protein and solvent, respectively.

Figure S1 shows that the main deviation between the results of MD simulation and master equation concerns the overestimation of the amplitudes at both ends of the peptide. As the N-terminal is charged and both terminal groups are moving quite freely in the solvent, this effect is most likely caused by an underestimation of the energy dissipation into the solvent at the termini. Enhancing this rate by a factor 3 for both terminal residues, we find a clearly better agreement of master equation and MD simulation results. Since we mainly aim for a simple global model of energy transport, however, for the remainder of this work we restrict ourselves to a single overall solvent dissipation rate.

The α-helical protein villin headpiece (HP36) represents another molecular system whose energy flow is dominated by backbone transport. In contrast to previous work on HP36,[86] here the energy transport simulations were run at room temperature (300 K) rather than at 10 K. Moreover, we employed a different MD force field and heated

residue Leu16 via T-jump excitation rather than via a photoswitch (cf. Sec. 2.7.1). Nonetheless, quite similar to the results of Ref.,[86] the time evolution of MD energies of the four next residues at both sides of the heater shown in Fig. S4 clearly reveals the signatures of backbone transport, that is, sequential energy propagation along the protein backbone. We note that the heat signal can be clearly detected up to 6 residues away from the heater. Moreover, we find slightly enhanced signals at residues Asp4 and Gln26, which obtain energy via contacts (4,15) and (18,26), respectively. While at 10 K a number of interresidue contacts were identified,[86] at 300 K only these two contacts were found to be structurally stable and at the same time relevant for energy transport.

Applying the above described parameterization procedure, Fig. S4 shows that the energy transport in HP36 is well described by a heating rate $k_\mathrm{h} = 1/3.1\,\mathrm{ps}$, a backbone diffusion coefficient $D_\mathrm{B} = 1.1\,\mathrm{nm}^2/\mathrm{ps}$, solvent dissipation rates $k_\mathrm{ps} = 1/8.8\,\mathrm{ps}$ and $k_\mathrm{sp} = 1/120\,\mathrm{ps}$, and contact rates $k_{4,15} = 1/9.2\,\mathrm{ps}$ (a triple contact evolving from a strong salt bridge between Asp4 and Arg15) and $k_{18,26} = 1/14\,\mathrm{ps}$ (a less stable hydrogen bond between Phe18 and Gln26). While contact rates and the heating rate depends on the specific system and the chosen heater residue, respectively, the more generic quantities values for k_ps, k_sp and D_B are quite similar as for Ala12, see Table 2.2. In line with a previous study,[41] $D_\mathrm{B} = 1.1\,\mathrm{nm}^2/\mathrm{ps}$ appears to be the same for extended structures and for α-helical structures, meaning that the hydrogen bonds of the α-helix do not significantly contribute to the energy transfer. Moreover, this choice of D_B is quite close to the value of $D_\mathrm{B} = 1.25\,\mathrm{nm}^2/\mathrm{ps}$ found by Buchenberg et al.[86] As the latter study was conducted at 10 K rather than 300 K, the backbone diffusion coefficient appears to depend only little on temperature.

2.8.2 Contact transport across a β-hairpin

To study the effects of interresidue contacts into the master equation, we first consider "AlaZip" (i.e., Ala12 equilibrated in a zipper structure) as a simple model of a β-hairpin, see Fig. 2.1. Although the system may become unstable at longer times (see Sec. 2.7.1), it remains in a stable β-hairpin structure during the 50 ps long nonequilibrium MD runs. Figure 2.10 shows the resulting time evolution of the residue energies $E_j(t)$. Since the initial structures are not completely equilibrated for stability reasons, most residue energies start somewhat above their equilibrium value. Compared to the results for Ala12 in Fig. S1, we moreover find significant enhancement of the residue energies at large sequence distances from the heater residue Val3 (up to a factor 4 at Ala12), which clearly indicates energy transport via interresidue contacts.

Most prominently, AlaZip features double interstrand hydrogen bonds between residue pairs (3,10) and (5,8), as well as a single hydrogen bond between residue pair (1,12), see Fig. 2.1). As a simple approximation for a master equation model for AlaZip, we therefore use identical backbone and solvent parameters (k_h, D_B, k_ps, k_sp) as for Ala12, and fitted the contact transport rates as $k_{1,12} = 1/6.3\,\mathrm{ps}$, $k_{3,10} = 1/2.0\,\mathrm{ps}$ and $k_{5,8} = 1/3.6\,\mathrm{ps}$.

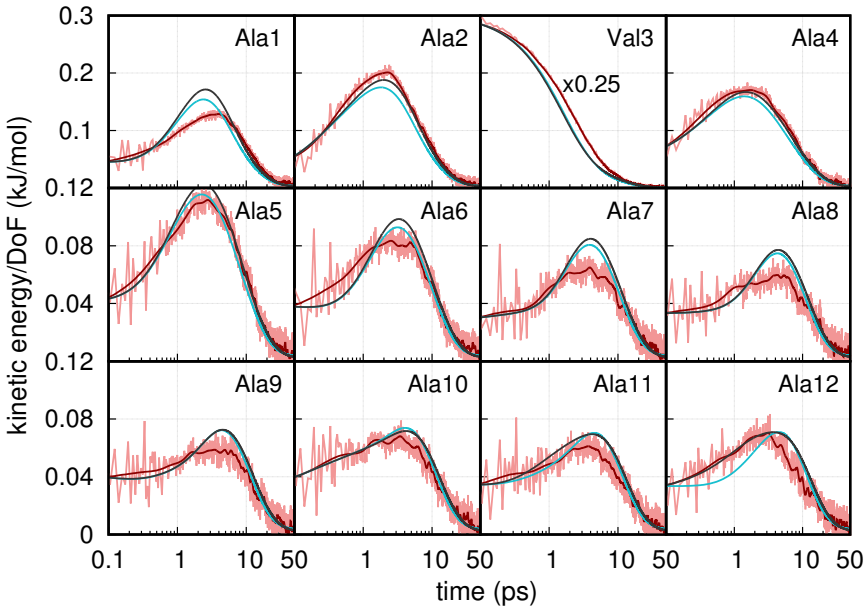


FIGURE 2.10: Time evolution of residue energies of AlaZip, obtained from nonequilibrium MD simulations (raw data in light red, Gaussian smoothed data in dark red) and from the master equation models with (black) and without (blue) heater contacts.

These findings indicate that the two double contacts (3,10) and (5,8) are about twice as effective than the single contact (1,12). While the rates associated with contacts (1,12) and (3,10) are defined with relatively small uncertainty (see below), the dominant parallel backbone transport between residues 5 and 8 renders the choice of $k_{5,8}$ somewhat ambiguous (say, $0 \leq k_{5,8} \leq 1/2\,\mathrm{ps}$). Figure 2.10 reveals that the resulting residue energies reproduce the MD data quite well.

However, already for the seemingly simple system AlaZip, the situation is more complicated. As a minor point, we first note that β-hairpin AlaZip exhibits a more compact structure than the (mostly) extended system Ala12, which results in a somewhat slower energy dissipation into the solvent ($1/k_{\mathrm{ps}} = 8.7\,\mathrm{ps}$ and $1/k_{\mathrm{sp}} = 250\,\mathrm{ps}$). More importantly, a closer structural analysis of AlaZip reveals weak nonpolar contacts between the side-chains of heater residue Val3 and residues Ala10 and Ala12 (Fig. 2.1). While these contact are significantly weaker than the interstrand contacts discussed above, the large energy content of the heater residue may render heater contacts quite important for the energy transport. To minimize the ambiguity associated with fitting of competing contact rates, we first estimated their relative contribution (see Appendix A.1). This yields for contact (1,12) the interstrand hydrogen bond contact rate $k_{1,12} = 1/17\,\mathrm{ps}$ and the heater contact rate $k_{0,12} = 1/21\,\mathrm{ps}$, as well as for contact (3,10) the rates $k_{3,10} = 1/10\,\mathrm{ps}$ and $k_{0,10} = 1/31\,\mathrm{ps}$. Since the rate of contact (5,8) is not well defined, we adopted a value of $k_{5,8} = 1/13\,\mathrm{ps}$ which is in line with the squared distance scaling rule discussed below. Figure 2.10 reveals that the resulting more realistic model reproduces the MD result with similar quality than the simpler model neglecting

heater contacts. As discussed in Sec. 2.8.3, though, the more realistic model allows us to parameterize scaling rules for hydrogen bond contact rates.

Let us now turn to the modeling of the energy transport in β-hairpin TrpZip2. Considering four mutations of TrpZip2 with different positions of Azu heater and Aha probing residues, upcoming experiments of Bredenbeck and coworkers have mapped out the energy flow in TrpZip2 with high spatial and temporal resolution.[39] We first focus on a TrpZip2 mutation with the Azu heater positioned in residue 3 (Fig. 2.1), which was recently studied using nonequilibrium MD simulations.[83] Similar to AlaZip, the system exhibits hydrogen bond contacts between residues pairs (1,12), (3,10), and (5,8), as well as complicated contacts of residues 10 and 12 with the aromatic ring of the Azu heater. The time evolution of MD residue energies shown in Fig. 2.11 again reveals signatures of backbone transport for the adjacent residues on each side of the heater, Trp2 and Ser1 as well as Trp4, Glu5 and Asn6. Beginning with Gly7 at the turn of the β-hairpin, effects of contact transport arise which are most prominent for the last three residues of the hairpin. For example, we find the earliest and strongest energy signal of Lys12, followed by Aha10 and then by Trp11. Compared to AlaZip, the residue energies [Eq. (2.7)] of TrpZip2 are generally smaller, because the residues of TrpZip2 contain more degrees of freedom.

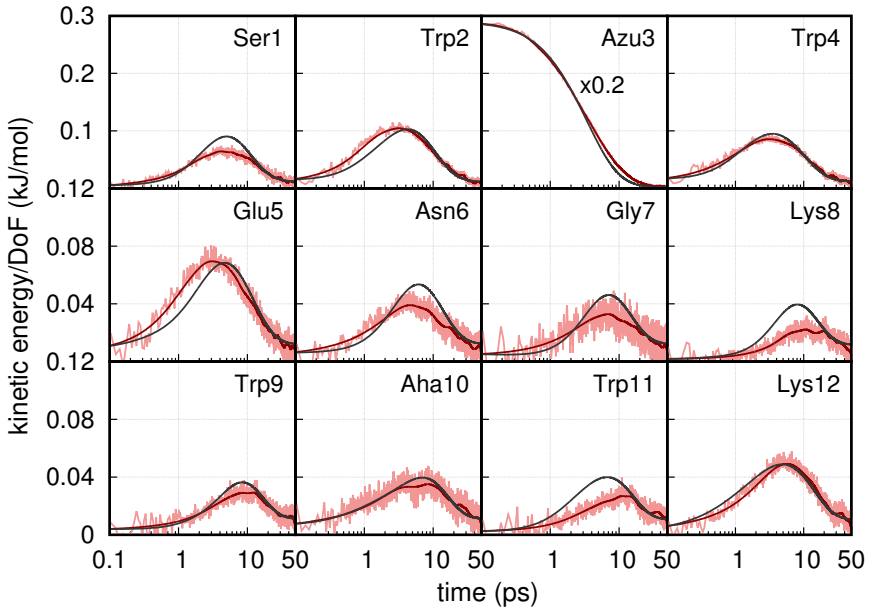


FIGURE 2.11: Time evolution of residue energies of TrpZip2, obtained from nonequilibrium MD simulations (raw data in light red, Gaussian smoothed data in dark red) and the from master equation model (black).

Applying our parameterization procedure, we obtain a heating rate $k_{\mathrm{h}} = 1/5.9\,\mathrm{ps}$, solvent dissipation rates $k_{\mathrm{ps}} = 1/8.3\,\mathrm{ps}$ and $k_{\mathrm{sp}} = 1/120\,\mathrm{ps}$, and a backbone diffusion coefficient $D_{\mathrm{B}} = 1.1\,\mathrm{nm}^2/\mathrm{ps}$. The contact rates are $k_{0,12} = 1/43\,\mathrm{ps}$, $k_{1,12} = 1/7.7\,\mathrm{ps}$, $k_{0,10} = 1/130\,\mathrm{ps}$ $k_{3,10} = 1/5.9\,\mathrm{ps}$, $k_{5,8} = 1/9.1\,\mathrm{ps}$ and $k_{0,5} = 1/71\,\mathrm{ps}$.. Figure 2.11 shows the resulting fit of the residue energies, which overall reproduces the MD results well.

In a similar way, we also modeled the energy transport for the other three mutants of TrpZip2, see Fig. S5 showing the fitted residue energies and Table S2 comprising all parameters.

2.8.3 Scaling Rules for contact transport

Having successfully described the energy flow in seven model peptides, we are now in a position to establish a general master equation model of energy transport. To begin, all systems can be described by the same backbone diffusion coefficient ($D_B = 1.1\,\mathrm{nm}^2/\mathrm{ps}$), a heating rate that reflects the size of the respective heater side-chain (e.g., $k_h = 1/5.9\,\mathrm{ps}$ for Azu), and a typical cooling time of $1/k_{ps} \approx$ 8-9 ps, see Table 2.2. Concerning contact transport, we have found complex nonpolar contacts with the side-chain of the heater, hence called "heater contacts". As shown by Gulzar et al.,[83] these interactions typically involves a number of atoms (e.g., of an aromatic ring) and can be therefore not appropriately described by a single distance. On the other hand, all β-hairpins exhibit intrastrand hydrogen bonds between residue pairs (1,12), (3,10) and (5,8). While the choice of contact rate $k_{5,8}$ turns out somewhat ambiguous due to dominant parallel backbone transport between residues 5 and 8, the rates associated with contacts (1,12) and (3,10) are defined with relatively small uncertainty. Hence these data are well suited to test the applicability of the above introduced scaling rules (2.9) and (2.10), which relate the rates to the variance $\langle \delta q_{ij}^2 \rangle$ and the squared distance $\langle q_{ij}^2 \rangle$ of the corresponding contact, respectively.

Depicting these rates as a function of $\sqrt{f_j/f_i}/\langle \delta q_{ij}^2 \rangle$ and $\sqrt{f_j/f_i}/\langle q_{ij}^2 \rangle$, Fig. 2.12 demonstrates the performance of the two scaling rules. The bars reflect the uncertainty of the rates, which were estimated from the quality decrease of the resulting residue energy fit when the rates were changed from their optimal value. Scaling rule (2.9) involving the contact distance variance $\langle \delta q_{ij}^2 \rangle$ is seen to clearly distinguish between (1,12) and (3,10) contacts. Fitting these cases separately, we obtain energy transport coefficients B_C of 3.5 and $1.1 \cdot 10^{-5}\,\mathrm{nm}^2/\mathrm{ps}$. As discussed in Sec. 2.7.3 and Appendix A.1, this reflects the fact that the rate (2.18) predicted by the model depend on the way the contact atoms are connected to the rest of the protein. That is, the rigidly connected contacts (3,10) involving two hydrogen bonds lead to smaller rates than the more loosely bound contacts (1,12) with a single hydrogen bond. This reasoning may also explain the findings of Reid et al.,[138] who observed that the hydrogen bond contact rates of myoglobin seem to fall on two lines. We note that the coefficient B_C found by Buchenberg et al.[86] for energy transport at 10 K is a factor 3-10 larger than our results obtained at 300 K, which reflects a decrease of the contact transport efficiency with increasing temperature.

The new scaling rule (2.10) depicted in Fig. 2.12b, on the other hand, is seen to describe all available contact rates with a single contact diffusion constant, $D_C = 2.1 \cdot 10^{-3}\,\mathrm{nm}^2/\mathrm{ps}$. This virtue make this scaling rule a versatile tool to predict the contact rates of proteins just by knowing the equilibrium distance between the contact atoms.

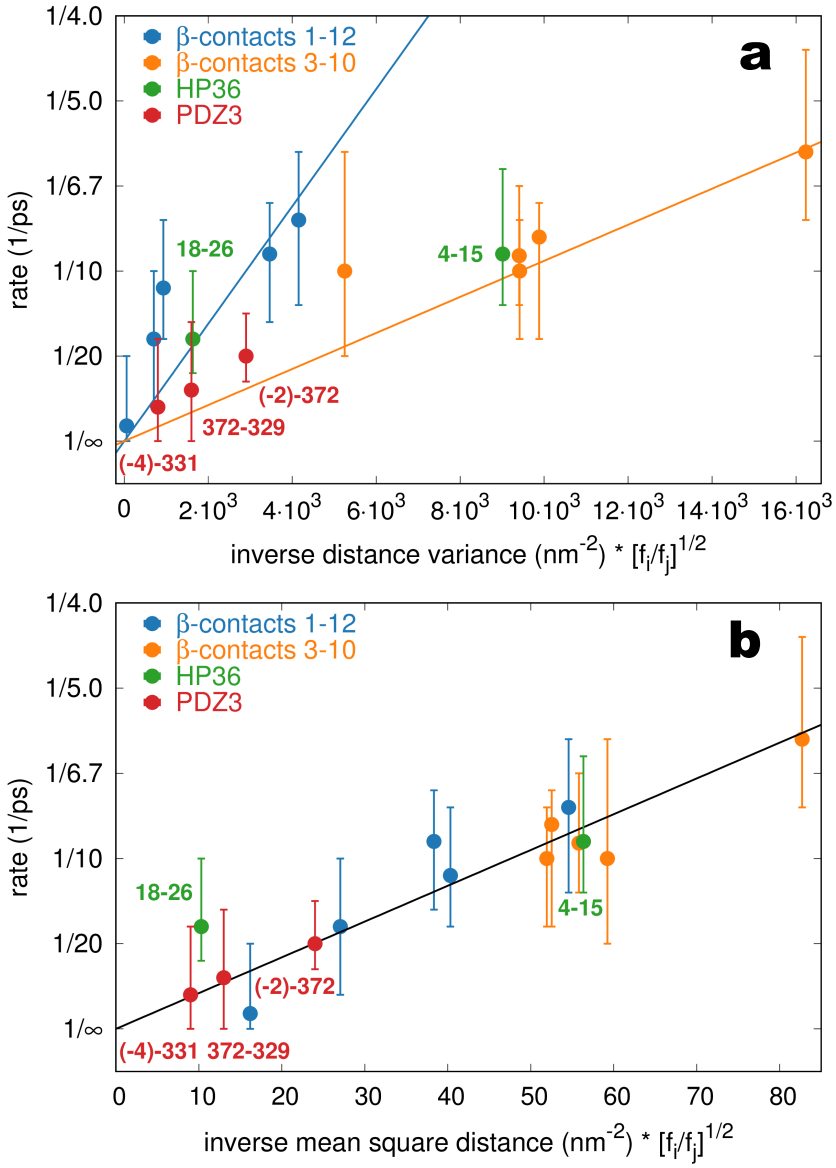


FIGURE 2.12: Scaling rules of contact transport. Shown are contact rates for interstrand hydrogen bonds (1,12) and (3,10) of the five studied β-hairpins (AlaZip, and mutations M1-M4 of TrpZip2), as well as most relevant contact rates found for HP36 and PDZ3. (a) Following scaling rule (2.9), the rates are plotted as a function of the inverse variance of the contact distance. Blue and orange lines represent linear fits ($R^2 = 0.45$ and 0.43) of the rates to Eq. (2.9) for (1,12) and (3,10) contacts, respectively. (b) Plotting the rates with respect to the inverse squared contact distance, the black line represents a fit ($R^2 = 0.84$) of all rates to Eq. (2.10).

2.8.4 Pathways of energy transport

Because the energy may flow through the backbone as well as via interresidue contacts, there is the question on the relevant energy transport pathways from an initially excited residue i to some target residue j. If the energy transport is modeled by a master equation, this question can be readily answered. Since the time evolution of a master

equation is completely determined by the rate matrix $\{k_{ij}\}$, we may run Monte Carlo Markov chain simulations which sample all possible energy transport pathways of the system with correct weights (see Sec. 2.7.5).

Adopting AlaZip and TrpZip2 as examples, we wish to study which pathways contribute when we excite residue 3 and probe residue 10 at the opposite side of the β-hairpin (Fig. 2.1). To this end, Fig. 2.13 shows the time evolution of the fraction of energy that a specific pathway contributes to the energy flow from residue 3 to residue 10. In the case of AlaZip, these residues are directly connected by two interstrand hydrogen bonds ($1/k_{3,10} = 10\,\mathrm{ps}$) and a heater contact ($1/k_{0,10} = 21\,\mathrm{ps}$). Nonetheless, we find that backbone represents the most effective transport channel with about 43 % of the total transported energy, while the hydrogen bond and the heater contact contribute 13 % and 23 %, respectively. That is, due to its high initial energy, the heater contact transports about twice the energy carried by the hydrogen bond, although its rate is only half the rate of the latter. The remaining 21 % of the energy is carried by various combinations of backbone and contact transport, e.g., a heater contact to Ala12 and subsequent backbone transport to Ala10 carries about 13 % of the energy. As may be expected, direct contacts (i.e., a one-step process) deliver the energy earlier than the detour through the backbone. The situation is somewhat more complex for TrpZip2,

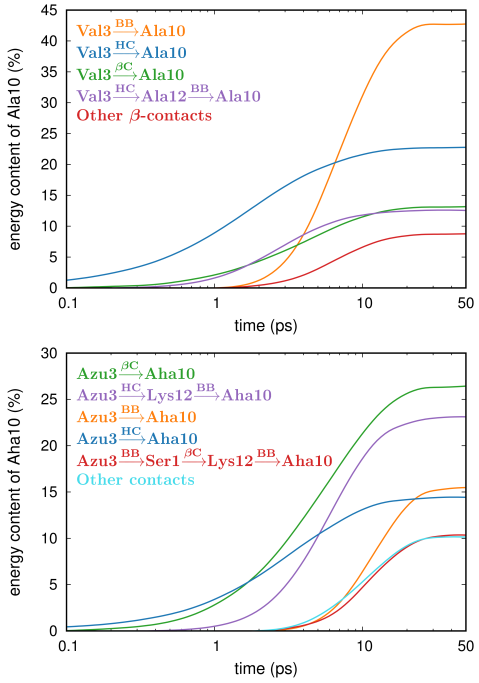


FIGURE 2.13: Contribution of various pathways to the energy transport from residue 3 to residue 10 of AlaZip (top) and TrpZip2 (bottom). Pathways may include the backbone (BB), interstrand hydrogen bonds (βC), heater contacts (HC), and other polar contacts (PC).

which contains diverse residues with in part large side chains. We find strong interstrand hydrogen bonds ($1/k_{3,10} = 5.9\,\mathrm{ps}$) that represent the most effective pathway

(27 %), followed by two heater contact pathways, either directly ($1/k_{0,10} = 125\,\mathrm{ps}$, 14 %) or via Lys12 ($1/k_{0,12} = 43\,\mathrm{ps}$, 23 %). Backbone transport contributes only 15 % of the energy. This is because several backbone transport steps are relatively slow, e.g., $\sim 1.5\,\mathrm{ps}$ for Trp4→Glu5, Lys8→Trp9 and Trp9→Aha10. This is a consequence of the large size of the involved residues and the associated long interatomic distances [cf. Eq. (2.5)]. Moreover, the succession of a large and a small residue causes a "bottleneck" for the energy flow, which is described in Eq. (2.5) by the detailed balance factor $\sqrt{f_j/f_i}$.

2.8.5 Anisotropic energy flow in PDZ3

To study the applicability of the above derived master equation model as well as the transferability of the scaling rules, we finally consider the energy transport in allosteric protein PDZ3,[128–130] for which first experimental results[82] as well as detailed nonequilibrium MD data are available.[83] The latter study revealed that PDZ3 exhibits a conformational transition in the binding pocket region, which may significantly affect the energy transfer of the system. Since a recent NMR study[155] indicated the population of a single conformation with the Azu residue turning to the β_2-sheet, here we restrict the analysis to this state.

Figure 2.14 shows the energy flow in PDZ3 as predicted by the MD simulations.[83] Since the Azu heater is attached to the N-terminus of the ligand, the energy propagates through the ligand backbone and successively reaches all its residues within a few picoseconds. Subsequently, the energy transfers via non-covalent contacts to the protein, where it reaches the α_2-helix and the β_2-sheet, as well as a few more distant residues. On the α_2 side, the ligand may form hydrogen bonds with His372, which is seen to receive some energy. On the β_2 side, residues 325 - 329 directly face the ligand and form contacts at positions 325 - 327 and 329, while Ile328 may form a side-chain contact directly with the Azu heater. Accordingly, residues 327 - 329 closest to the heater are seen to obtain more energy and faster than residues 325 and 326. There are also a few contacts to neighboring residues that are more remote in sequence space. In particular, Glu331 and Phe400 receive energy via heater contacts. Following the procedure outlined above, we use the standard value of the backbone diffusion coefficient ($D_\mathrm{B} = 1.1\,\mathrm{nm}^2/\mathrm{ps}$), the heating rate of Azu ($k_\mathrm{h} = 1/5.9\,\mathrm{ps}$), and determined solvent dissipation rates $k_\mathrm{ps} = 1/10\,\mathrm{ps}$ and $k_\mathrm{sp} = 1/60\,\mathrm{ps}$. The somewhat larger overall cooling time of 10 ps reflects the fact that, compared to the smaller systems studied above, the residues of PDZ3 are only in part exposed to the solvent. Concerning contact transport, only contacts close to the Azu heater need to be taken into account. We found 9 relevant interstrand hydrogen bonds pertaining to the β-sheets of PDZ3, whose rates were obtained using scaling rule (2.10), see Table S3. Moreover, we identified 4 heater contacts with surrounding residues, 2 hydrogen bonds between ligand and β_2 sheet, a hydrogen bond between Gly329 and His372, as well as a nonpolar contact involving the beta-CH2

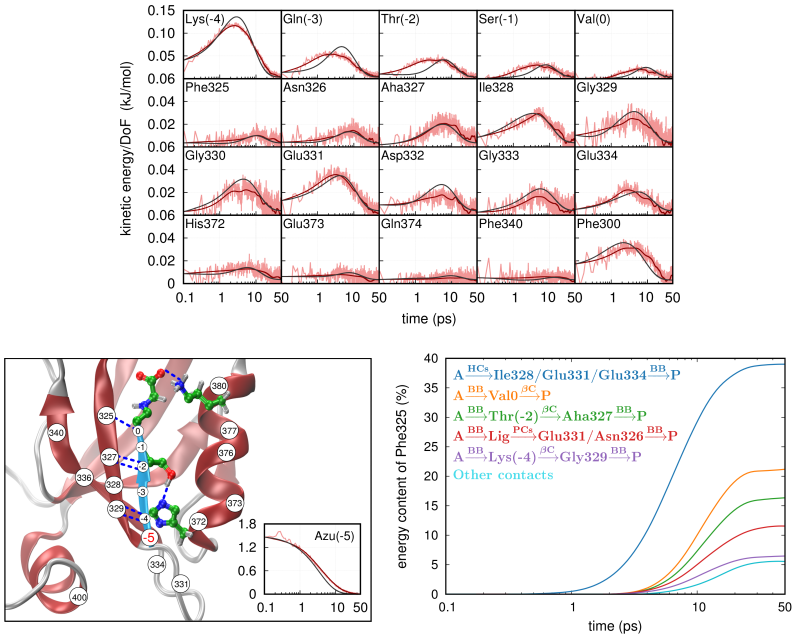


FIGURE 2.14: Energy transfer in PDZ3. Top: Time evolution of residue energies, reflecting the energy flow from the ligand via α_2 helix and β_2-sheet to more remote regions. Bottom left: Close-up view of the binding pocket, including the ligand (green) and the adjacent α_2 and β_2 regions of the protein (red). By convention, the ligand is labelled from -5 to 0, while the protein is labelled from 300 to 415. Bottom right: Main energy transport pathways from the initially excited ligand residue Azu(-5) (A) to residue Phe 325 (P). Pathways may include the backbone (BB), interstrand hydrogen bonds (βC), heater contacts (HC), and other polar contacts (PC).

groups of residues Lys(-4) and His372, see Table 2.2 for the associated rates. Employing these rates to test the applicability of scaling rules (2.9) and (2.10), Fig. 2.12 shows that hydrogen bond contacts are mostly well described by both scaling rules. The exception is the double hydrogen bond between Thr(-2) and His372, which might be related to the fact that His372 may also form several partly populated other contacts (particularly with Lys(-4)) that are difficult to describe. Figure 2.14 shows the resulting master equation prediction of the residue energies, which reproduces the MD results very well.

As an example of energy transport pathways, we finally consider the transport from the initially excited residue of the ligand Azu(-5) to the 1.7 nm distant residue Phe325. Figure 2.14 demonstrates that a number of different combinations of backbone and contact transport steps exist. Most effective are direct heater contacts with either Ile328, Glu331 or Glu334 followed by backbone transport (39 %), as well as a pathway featuring backbone transport to Val(0) at the end of the ligand, followed by a direct hydrogen bond contact to Phe 325 (21 %). Furthermore, various more complicated combinations backbone and contact transport are found.

2.9 Conclusions

We have outlined a general master equation model that describes the energy transport in proteins. It relies on scaling rules (2.5) and (2.10) that predict the rates of backbone and contact transport solely from the molecular structure of the system. Despite of its deliberate simplicity that neglects many details shown by explicit MD simulations, the model was found to describe the energy flow in various systems qualitatively correctly. Moreover, the master equation model allows us to map out the individual pathways of energy transport and assess their importance. The construction and performance of the model highlighted the following issues of protein energy transport.

1. Energy transport in solvated proteins is limited by dissipation of the energy into the surrounding solvent. In good agreement of experiment[44,82] and simulation,[83,131] this cooling of the protein in water occurs on a timescale of 8 - 10 ps at 300 K (Table 2.2). While more sophisticated models can be considered, we have chosen to adopt a common cooling rate for all residues of the system.

2. In all systems considered, the transport along the protein backbone represents the fastest channel of energy flow. It is well described by a diffusive model giving rise to scaling rule (2.5), which requires only an overall backbone diffusion coefficient ($D_B = 1.1\,\mathrm{nm}^2/\mathrm{ps}$) and interatom distances as input. Typical transfer times between adjacent residues are 0.5 - 1 ps. In our classical model, backbone transport was found to depend only little on temperature.

3. Transport through interresidue contacts, on the other hand, was shown to depend strongly on temperature. This is because thermal fluctuations may increase the flexibility of the atom groups forming the contact, which weakens the connection. While previous MD studies conducted at low temperatures predicted effective contact transport, our room temperature study indicates less importance of interresidue contacts. That is, we have found typical transfer times of 6 - 30 ps (Table 2.2), which renders contact transport at least one order of magnitude slower than backbone transport.

4. The previously suggested inverse variance rule[86] of contact transport was shown to depend on the way the contact atoms are connected to the rest of the protein (see Appendix A.1), which limits the practical use of the theory. The new inverse square distance scaling rule (2.10), on the other hand, was found to describe virtually all available contact rates with a single contact diffusion constant, $D_C = 2.1 \cdot 10^{-3}\,\mathrm{nm}^2/\mathrm{ps}$. To establish the general validity of this scaling rule, clearly more systems need to be considered.

5. We have shown that typically only polar contacts are relevant for energy transport. Even nonpolar contacts arising from prominently stacked aromatic rings as

in HP36 and TrpZip2 are found to hardly contribute to the energy flow. While these mostly entropic interactions may be crucial to stabilize the structure of a protein, they result in comparatively small forces at equilibrium distance, which is what's needed for energy transport. In this sense, the energy flow in a protein resembles its interresidue force network, which has been employed to explain allosteric communication in proteins.[163]

6. Employing Monte Carlo Markov chain simulations of the master equation, we have identified the relevant energy transport pathways of the considered system. As shown by the comparison of two seemingly similar systems, AlaZip and TrpZip2, the competition between backbone and contact transport may depend on details of the protein sequence and structure.

The description of energy transport experiments requires the inclusion of heater residues, which may form efficient contacts with adjacent residues that complicate matters. This is not an issue, if we do not intend to simulate a specific experiment, but rather are interested in the general flow of vibrational energy through a given protein. In this case the above described general master equation model based on the simple scaling rules (2.5) and (2.10) should be sufficient to make qualitative predictions of the protein energy flow. Since the scaling rules require only the structure of the system under consideration, the model provides a simple and general means to predict the energy transport pathways in proteins.

2.10 Supplementary Information

Supplementary Methods: Relative contributions of heater and hydrogen bonds contacts

To calculate the energy transport contributions of a hydrogen bond contact and a heater contact connecting residue $j = 1$ and heater residue h, we rewrite master equation (1) such that the contributions due to these two contacts are taken out of the sum

$$\frac{dE_1}{dt} = \sum_i [k_{i,1}E_i - k_{1,i}E_1] + k_{\mathrm{S},1}E_\mathrm{S} - k_{1,\mathrm{S}}E_1 + k_{\mathrm{B},1}E_\mathrm{B} - k_{1,\mathrm{B}}E_1. \tag{2.12}$$

Here E_S and E_B are the kinetic energies of side-chain and backbone atoms of the heater residue, respectively, and $k_{\mathrm{B},1}$ and $k_{\mathrm{S},1}$ are the contact transport rates from these atom groups to residue 1. Integration from time t_0 to t_f yields

$$\frac{\Delta E_1}{t_f - t_0} = \sum_i [k_{i,1}\bar{E}_i - k_{1,i}\bar{E}_1] + k_{\mathrm{S},1}(\bar{E}_\mathrm{S} - f_1/f_\mathrm{S}\bar{E}_1) + k_{\mathrm{B},1}(\bar{E}_\mathrm{B} - f_1/f_\mathrm{B}\bar{E}_1), \tag{2.13}$$

where $\Delta E_1 = E_1(t_f) - E_1(t_0)$, $\bar{E}_j = 1/(t_f - t_0)\int_{t_0}^{t_f} E_j(t)dt$, and we used detailed balance condition (see Eq. (3) in main text). When we assume that all other rates $k_{i,1}$ are known and calculate energies $\bar{E}_j$ from nonequilibrium MD simulations (using $t_0 = 0.1\,\mathrm{ps}$ and $t_f = 10\,\mathrm{ps}$), we recast this into

$$k_{\mathrm{S},1} = a_1 - b_1 k_{\mathrm{B},1}, \tag{2.14}$$

describing the relative contributions of heater contact and hydrogen bond contact to the energy transport to residue 1. Using this relation, the rates of the two competing contacts can be readily obtained by minimizing the RMSD. Table S3 list the resulting parameters a_i and b_i obtained for all respective residues of AlaZip and the four mutations of TripZip2, together with the resulting heater contact and hydrogen bond contact rates.

System	AlaZip		TrpZipM1		TrpZipM2		TrpZipM3		TrpZipM4	
Contact	(1,12)	(3,10)	(1,12)	(3,10)	(1,12)	(3,10)	(1,12)	(3,10)	(1,12)	(3,10)
a (1/ps)	0.035	0.068	-	-	0.024	0.018	0.013	-	0.011	-
b	0.044	0.21	-	-	0.0039	0.059	0.090	-	0.084	-
$1/k_{\mathrm{h}i,j}$ (ps)	31	21	-	-	43	130	300	-	300	-
$1/k_{i,j}$ (ps)	17	10	111	9.2	7.7	5.9	9.1	8.3	11	10
$\sqrt{f_i/f_j}/\langle\delta q_{ij}^2\rangle$ (1/nm^2)	702	5249	55	9407	4152	16234	3463	9881	930	9416
$\sqrt{f_i/f_j}/\langle q_{ij}^2\rangle$ (1/nm^2)	27	59	16	56	55	83	38	53	40	52

TABLE S3: Relative contributions of competing heater and hydrogen bonds contacts of AlaZip and the four mutations of TripZip2, see the Appendix of the main paper. Shown are the parameters a_i and b_i of Eq. (2.14), the resulting heater contact rates $k_{\mathrm{h}i,j}$ and hydrogen bond contact rates $k_{i,j}$, as well as the squared contact distances $\langle q_{ij}^2\rangle$ and variances $\langle\delta q_{ij}^2\rangle$ of all considered contacts multiplied by the corresponding factor $\sqrt{f_i/f_j}$.

Supplementary Results

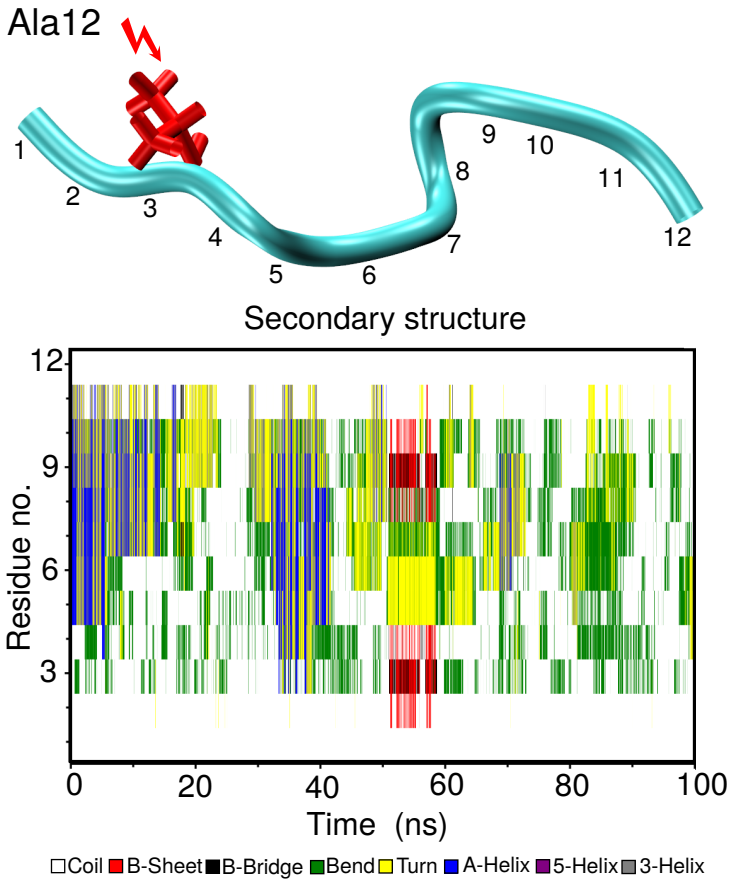


FIGURE S1: Structure (top) and configurations (bottom) of Ala12 during a 100ns long equilibrium simulation performed as described in Sec. II A. In red, frames where the residues exhibit a β sheet conformation.

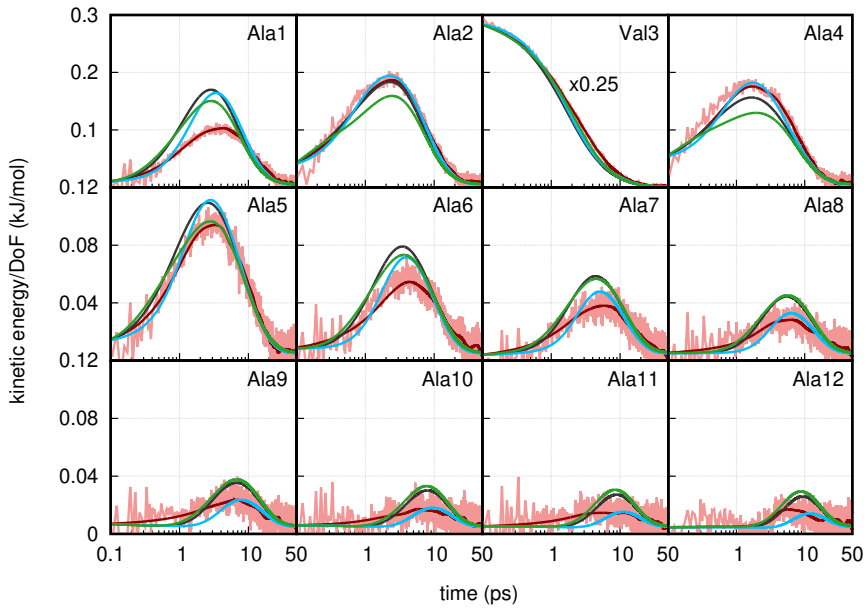


FIGURE S2: Sensitivity of master equation model with respect to the choice of heater rate k_h and the backbone diffusion coefficient. Compared are results for residue energies of Ala12, obtained from nonequilibrium MD simulations (raw data in light red, Gaussian smoothed data in dark red) and master equation models using standard values ($1/k_h = 1.7\,\text{ps}$ and $D_B = 1.1\,\text{nm}^2/\text{ps}$, black lines) and alternative choices ($1/k_h = 1.3\,\text{ps}$ and $D_B = 0.6\,\text{nm}^2/\text{ps}$, blue lines) and ($1/k_h = 2.5\,\text{ps}$ and $D_B = 1.4\,\text{nm}^2/\text{ps}$, green lines).

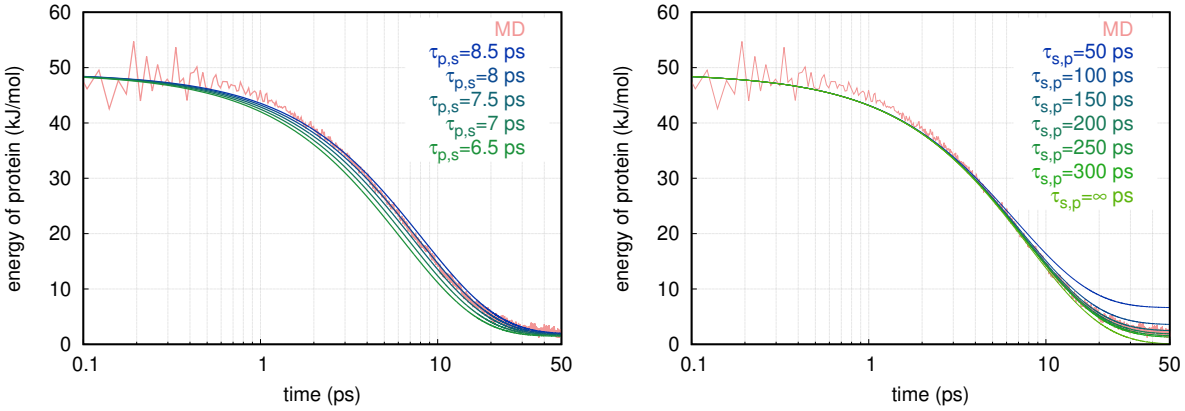


FIGURE S3: Cooling of the protein in the solvent; shown are fits of the total protein energy of Ala12 for $\tau_{sp} \equiv 1/k_{sp} = 210\text{ps}$ and several values of $\tau_{ps} \equiv 1/k_{ps}$ (left); and vice versa, for $\tau_{ps} = 7.9\text{psb}$ and different values of τ_{sp} (right).

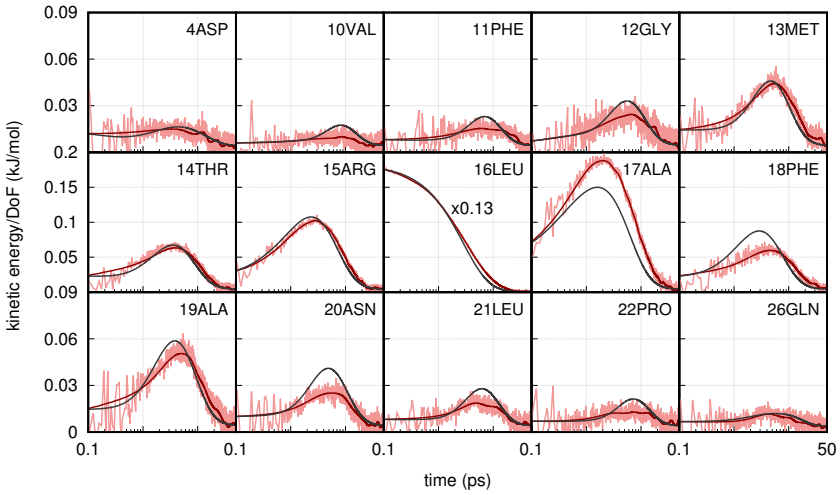


FIGURE S4: Time evolution of selected residue energies of HP36, following initial excitation of Leu16. Shown are results from nonequilibrium MD simulations (raw data in light red, Gaussian smoothed data in dark red) and from the master equation model (black).

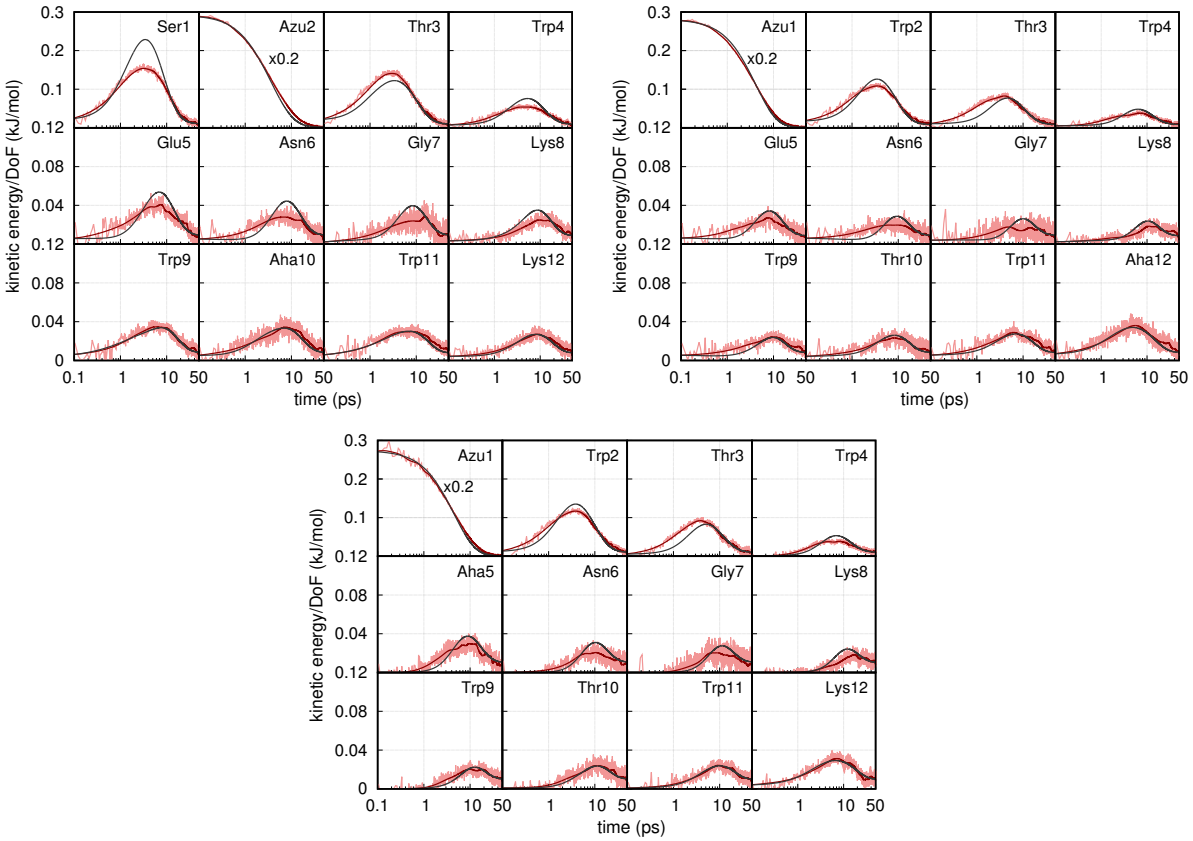


FIGURE S5: Time evolution of residue energies of mutants 1 (top left), 3 (top right) and 4 (bottom) of TrpZip2 with results from nonequilibrium MD simulations (raw data in light red, Gaussian smoothed data in dark red) and from the master equation model (black).

System	Mutant1	Mutant3	Mutant4
Heater	Azu2	Azu1	Azu1
$1/k_{\mathrm{h}}$	1.7	1.7	1.7
$1/k_{\mathrm{ps}}$	8.3	7.5	8.9
$1/k_{\mathrm{sp}}$	165	120	235
contact times	$\tau_{1,12} = 110$ $\tau_{3,10} = 9.2$ $\tau_{5,8} = 9.1$ $\tau_{0,9} = 59$ $\tau_{0,11} = 77$	$\tau_{1,12} = 9.1$ $\tau_{3,10} = 8.3$ $\tau_{5,8} = 9.1$ $\tau_{0,12} = 300$ $\tau_{0,11} = 300$	$\tau_{1,12} = 11$ $\tau_{3,10} = 10$ $\tau_{5,8} = 9.1$ $\tau_{0,12} = 300$

TABLE S4: List of considered mutations of TrpZip2, heater residues and resulting master equation parameters, including inverse heating rate $1/k_{\mathrm{h}}$, solvent cooling time $1/k_{\mathrm{ps}}$, solvent back-transfer time $1/k_{\mathrm{sp}}$ and contact times $\tau_{\mathrm{ij}} = 1/k_{\mathrm{ij}}$, all in units of ps.

Contacts	Lys(-4)→Gly329	Thr(-2)→Aha327	Val0→Phe325	Leu342→Gly324
$1/k_{\mathrm{ij}}$ (ps)	43	9.5	15	28
$\sqrt{f_i/f_j}/\langle q_{\mathrm{ij}}^2\rangle$ (1/nm^2)	11	50	31	17
$\sqrt{f_i/f_j}/\langle \delta q_{\mathrm{ij}}^2\rangle$ (1/nm^2)	58	5638	971	3266

Phe340→Asn326	Ser339→Asn326	Phe337→Ile328	Gly335→330	Glu334→Glu331
14	17	9.5	29	23
33	27	50	16	21
5485	3780	11354	142	1080

TABLE S5: Contact rates k_{ij} of the relevant hydrogen bonds of PDZ3 and their corresponding squared contact distances $\langle q_{\mathrm{ij}}^2\rangle$ and variances $\langle \delta q_{\mathrm{ij}}^2\rangle$ multiplied by the corresponding factor $\sqrt{f_i/f_j}$.

A.1 Harmonic model of contact transport

To describe the through-space energy transport via a contact between two protein residues, Fig. A.1 shows a one-dimensional harmonic model of a contact between atoms 1 and 2 with mass-weighted coordinates q_1 and q_2 and spring constant d. To the left and right the atoms are covalently bound (described by spring constants D_L and D_R) to adjacent atoms L and R, which themselves are bound to other atoms of the respective residue. The equation of motion for coordinate q_1 reads

$$\ddot{q}_1 = -d(q_1 - q_2) - D_L(q_1 - q_L), \tag{2.15}$$

and similarly for q_2. By solving these equations, our aim is to calculate the kinetic energies of atoms 1 and 2, in order to account for their interatom energy transport. To this end, we need to know the time evolution of coordinates q_L and q_R, whose equations of motion depend on the way atoms L and R are connected to the rest of the protein. To obtain a concrete model, we here assume that the overall structure of the protein is quite rigid, such that atoms L and R are approximately fixed (i.e., $q_L = q_R = 0$). For simplicity, we furthermore assume that $D_L = D_R = D$.

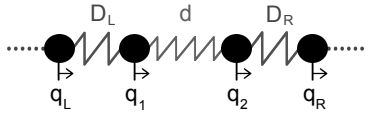


FIGURE A.1: Scheme of the one-dimensional harmonic model of contact energy transport.

By introducing coordinates $q = q_1 - q_2$ and $Q = q_1 + q_2$, the system decouples in equations

$$\ddot{q} + (D + 2d)q = 0, \quad \ddot{Q} + DQ = 0, \tag{2.16}$$

describing relative motion and center-of-mass motion with harmonic frequencies $\omega = \sqrt{D + 2d}$ and $\Omega = \sqrt{D}$, respectively. From the solution of Eq. (2.16), we readily obtain local modes $q_1 = (Q+q)/2$ and $q_2 = (Q-q)/2$.

To model the energy transfer between atoms 1 and 2, we assume that initially atom 1 is excited via $\dot{q}_1(0) = v_0$ and that $\dot{q}_2(0) = q_1(0) = q_2(0) = 0$. Hence the kinetic energy of atom 2 reads

$$E_2(t) = \frac{\dot{q}_2^2}{2} = v_0^2 \sin^2 \frac{(\omega + \Omega)t}{2} \sin^2 \frac{(\omega - \Omega)t}{2}. \tag{2.17}$$

Averaging over the rapidly oscillating term with $\omega + \Omega$, we focus on the slowly varying term $\propto \sin^2 \frac{(\Omega - \omega)t}{2}$, which reaches its maximum at $(\Omega - \omega)t = \pi$. If we define the energy rise time τ_{12} as half of this value, we obtain for the energy transfer rate $k_{12}^{\mathrm{C}} = 1/\tau_{12} = 2/\pi\ (\omega - \Omega) = 2/\pi\ (\sqrt{D + 2d} - \sqrt{D})$. Assuming that the force constant D of a covalent bond is significantly larger than the force constant d associated with a typical

hydrogen bond, we approximate $\sqrt{D+2d}-\sqrt{D} \approx d/\sqrt{D}$ and finally obtain

$$k_{12}^{\mathrm{C}} = \frac{2}{\pi}\frac{d}{\sqrt{D}}. \tag{2.18}$$

Since for a harmonic oscillator $\langle V\rangle = d/2$ $\langle \delta q^2\rangle = \langle E_{\mathrm{kin}}\rangle = k_{\mathrm{B}}T/2$, the spring constant d can be calculated from the variance of the displacement $\delta q = (q - \langle q\rangle)$ via $d = k_{\mathrm{B}}T/\langle \delta q^2\rangle$. Using $B_{\mathrm{C}} = 2k_{\mathrm{B}}T/(\pi\sqrt{D})$ then leads to scaling rule (2.9). Note that the model predicts that rigidly connected contacts (with a large spring constant D) lead to smaller rates than loosely bound contacts.

2.11 Comparison with experiments

To relate time-dependent IR experiments, nonequilibrium MD simulations and predictions of master equation we again considered four mutants of TrpZip2 as shown in Fig. 2.11 and S5. In Fig. 2.15 and Table 2.6 we compare experimental results with calculated residue energies of Aha in the various TrpZip2 mutations. The comparison assumes that the calculated vibrational excess energy of Aha coincides with the transient red shift ("hotband") of the azido stretch band of Aha.[87] Using azido stretch band as a local probe provides an additional advantage over previously employed amide I band[38,41] in that it exhibits similar time evolution of azido stretch energies as compared to Aha residue energies computed from nonequilibrium MD simulations. This might stem from the fact that azido mode represents an isolated local vibrational mode, whereas, amide I band reflects the coupling of many C=O oscillators. Overall, the MD calculations are found to overestimate the peak times of energy transport by about a factor 2, which is a consequence of the classical nature of MD simulations. Classical calculations neglect the quantum fluctuations such as vibrational zero point energy and underestimate vibrational energy transition rates due to high-frequency vibrations.[154,164] While energy redistribution via classical equilibrium calculations are often wrong by several orders of magnitude,[164] nonequilibrium calculations typically require quantum corrections of about a factor 2-3.[154]

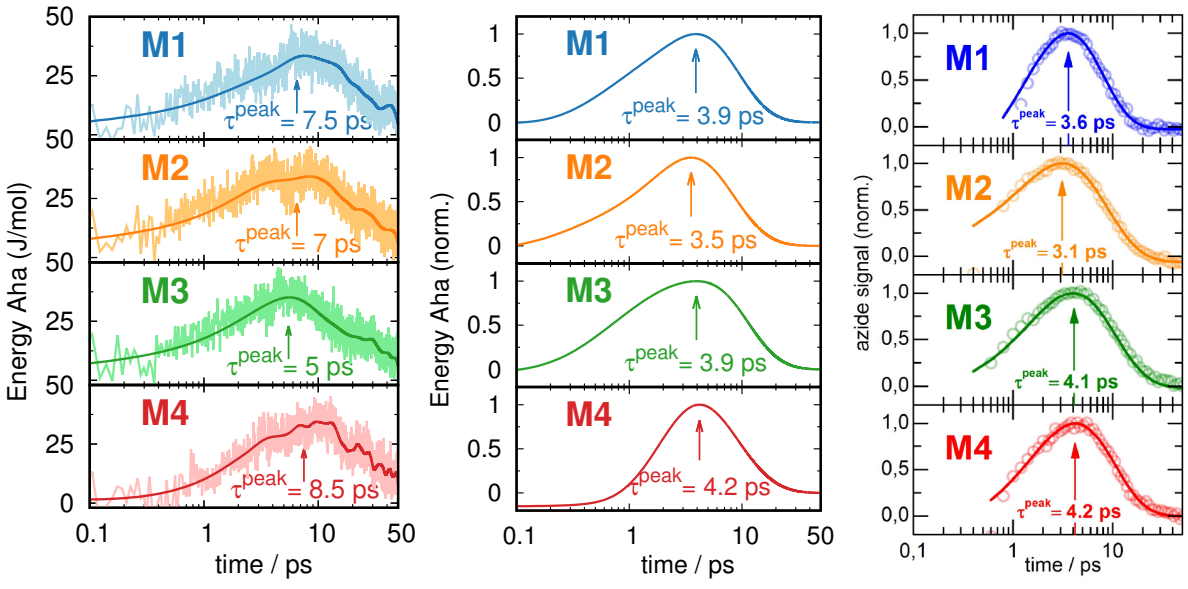


FIGURE 2.15: Evolution of Aha energies of M1-M4 of TrpZip2 and their respective peak-times; obtained directly from MD simulation (Left) and from master equation model (Middle) by scaling contact and backbone transport, as well as transport within the heater by a factor 3.1 (i.e. $D_B = 3.4\text{nm}^2\text{ps}^{-1}$, $D_C = 6.5 \cdot 10^{-3}\text{nm}^2\text{ps}^{-1}$ and $k_h = 1/1.9\text{ps}$). (Right). Experimental data shown prior to publication[39]

TABLE 2.6: Comparison of experimental and calculated solvent dissipation times τ^{solv}, peak times τ^{peak} and and decay times τ^{decay} (in ps), as exhibited by mutations M1 to M4 of TrpZip2.

System	Azu-Aha	$\tau^{\text{solv}}_{\text{exp}}$	$\tau^{\text{solv}}_{\text{MD}}$	$\tau^{\text{peak}}_{\text{exp}}$	$\tau^{\text{peak}}_{\text{MD}}$	$\tau^{\text{decay}}_{\text{exp}}$	$\tau^{\text{decay}}_{\text{MD}}$
M1	2-10	5.1	8.3	3.6	7.5	10	30
M2	3-10	5.4	8.3	3.1	7	10	32
M3	1-12	6.3	7.5	4.1	5	12	33
M4	1-5	6.4	8.9	5.5	8.5	12	32

In order to obtain a better agreement between predictions of master equations and experimental results we attempted to determine quantum corrections for the heater rate k_{h}, the backbone diffusion coefficient D_{B}, heater contact rates and the contact transport coefficient D_{C}. By directly using experimental solvent dissipation times $\tau^{\text{solv}}_{\text{exp}}$ we find that a global scaling factor 3.1 (see Fig. 2.15 (middle)) reproduces quite realistic peak- and decay-times of all four TrpZip2 mutants however, rising behaviors slightly differ from experimental results.

Chapter 3

Sensivity of AHA label

This chapter is based on following publications:

2D-IR Spectroscopy of an AHA Labelled Photoswitchable PDZ2 Domain

Brigitte Stucki-Buchli[a], Philip J. M. Johnson[a], Olga Bozovic[a], Claudio Zanobini[a], Klemens L. Koziol[a], Peter Hamm[a,] Adnan Gulzar[b], Steffen Wolf[b], Sebastian Buchenberg[b] and Gerhard Stock[b,*]*

J. Phys. Chem. A 2017, 121, 9435-9445

Azidohomoalanine: A Minimally Invasive, Versatile, and Sensitive Infrared Label in Proteins To Study Ligand Binding

Claudio Zanobini[a], Olga Bozovic[a], Brankica Jankovic[a], Klemens L. Koziol[a], Philip J. M. Johnson[a], Peter Hamm[a,], Adnan Gulzar[b], Steffen Wolf[b] and Gerhard Stock[b,*]*

J. Phys. Chem. B 2018, 122, 10118-10125

[a]Department of Chemistry, University of Zurich, Zurich, Switzerland
[b]Biomolecular Dynamics, Institute of Physics, Albert Ludwigs University, Freiburg, Germany
**corresponding authors: peter.hamm@chem.uzh.ch and stock@physik.uni-freiburg.de*

We explore the capability of the non-natural amino acid azidohomoalanine (AHA) as an IR label to sense relatively small structural changes in proteins with the help of 2D IR difference spectroscopy. To that end, we AHA-labelled an allosteric protein (the PDZ2 domain from human tyrosine-phosphatase 1E) and furthermore covalently linked it to an azobenzene-derived photoswitch as to mimic its conformational transition upon ligand binding. To determine the strengths and limitations of the AHA label, in total six mutants have been investigated with the label at sites with varying properties. Only one mutant revealed a measurable 2D IR difference signal. In contrast to the commonly observed frequency shifts which report on the degree of solvation, in this case we observe an *intensity* change. To understand this spectral response, we performed classical MD simulations, evaluating local contacts of the AHA labels to water molecules and protein side chains and calculating the vibrational frequency based on an electrostatic model. While these simulations revealed in part significant and complex changes of the number of intraprotein and water contacts upon *trans-cis* photoisomerization, they

could not provide a clear explanation of why this one label would stick out. Subsequent quantum-chemistry calculations suggest that the response is the result of an electronic interaction involving charge transfer of the azido group with sulphonate groups from the photoswitch. To the best of our knowledge, such an effect has not been described before.
To further investigate the environment-dependent sensitivity of AHA as an infrared probe for the site-specific investigation of protein structure and dynamics we explore its capability to detect protein-ligand interactions by incorporating it in the vicinity of the binding groove of a PDZ2 domain. Circular dichroism (CD) and isothermal titration calorimetry (ITC) measurements reveal that the perturbation of the protein system by the mutation is negligible, with minimal influence on protein stability and binding affinity. Two dimensional infrared (2D IR) spectra exhibit small (1-3 cm^{-1}) but clearly measurable red shifts of the Aha vibrational frequency upon binding of two different peptide ligands, while accompanying molecular dynamics (MD) simulations suggest that these red shifts are induced by polar contacts with side chains of the peptide ligands. Hence, Aha is a versatile and minimally invasive vibrational label that is not only able to report on large structural changes during e.g. protein folding, but also on very subtle changes of the electrostatic environment upon ligand binding.

2D-IR Spectroscopy of an AHA Labelled Photoswitchable PDZ2 Domain
Here, I performed all MD simulations and analyzed the simulation data.

3.1 AHA labelled photoswitchable PDZ2 Domain

Proteins are dynamical objects. The structural dynamics of proteins involve equilibrium processes, such as thermally driven fluctuation, as well as non-equilibrium processes, such as the conformational transition in a light-triggered protein. Vibrational spectroscopy provides an inherent picosecond time resolution to study both equilibrium and non-equilibrium processes. However, obtaining site-selective information from vibrational spectroscopy will in general require vibrational labels, since the IR spectrum of a molecule of that size is no longer resolved into its normal modes. Ideally, such vibrational labels should absorb outside of the congested region of the absorption spectrum of a protein[88–92] in order to discriminate it from a huge background. Various distinct molecular groups have been suggested for that purpose: –SH vibrations of cysteines,[165] –CD vibrations of deuterated amino acids,[166–168] –C≡O vibrations of metal-carbonyls either from natural cofactors such as a heme group[169–171] or from complexes that can be bound to amino acids in a post-translational step,[172–174] as well as $–N_3$,[71,93,94,175–178] –C≡N[179,180] and –SCN[181] vibrations from non-natural amino acids. All these molecular groups have in common that they absorb in a spectral window between ≈1700 cm^{-1} and ≈2800 cm^{-1}, where essentially no fundamental modes of natural proteins are found

(with the one exception of the –SH vibrations of cysteines), and where the only background originates from the still quite strong but very broad and featureless absorption of the solvent water.

In addition to having a frequency in that spectral window, a good label should fulfill the following criteria:

- The extinction coefficient of the label should be large enough so that it can be detected at reasonable concentrations. As most proteins are not soluble at high concentrations, the goal is to measure at concentrations of around 1 mM or below. This concentration range is comparable to what is commonly used for NMR measurements.
- It should be versatile and incorporable at essentially any position of a protein with good yields and at a high purity (since IR measurements require large amounts of sample).
- It should not significantly perturb the structure and the stability of a protein.
- The label should be sensitive to its environment, e.g., sense the polarity or hydrophobicity of its surrounding.

With these criteria in mind, we currently concentrate on the non-natural amino acid azidohomoalanine (AHA) as a label, which contains an azido group ($-N_3$) that absorbs at around 2100 cm^{-1}. It has a reasonably high extinction coefficient of 300-400 $M^{-1}cm^{-1}$,[93] large enough to be measured at concentrations well in the sub millimolar regime by 2D IR spectroscopy.[91,94] As a small amino acid that is a methionine analog, it can be incorporated also into larger proteins (which can no longer be synthesized on a peptide synthesizer) at essentially any position by a methionine auxotrophic protein expression strategy.[182,183] It does not perturb protein properties very much, as evidenced for example by the fact that labelling a peptide ligand with AHA affects its binding affinity to a PDZ2 domain only to a small extent.[93] Finally, AHA has been shown to be a sensitive probe of its environment. For example, when AHA is buried in the hydrophobic core in the folded state,[176] the azido band blue-shifts by up to 20 cm^{-1} upon protein unfolding. Along the same lines, it can also be used to detect binding of an AHA-labelled peptide ligand to a larger protein, in which case the degree of solvation of the label diminishes, causing a red-shift of its vibration.[93,94] Such shifts have been explained mainly by changes in hydrogen bonding to the first and/or third nitrogen atom of the azido group as well as to variations in the angle between the hydrogen bond and the azido group.[70,178] Furthermore, the vibrational frequency is also influenced by the polarity of the environment,[178] e.g., Coulomb interactions to the middle nitrogen atom of the azido group can shift the vibration.[71]

In the present study, we explore the capability of the AHA label to sense relatively small structural changes of a protein, i.e., changes much smaller than those occurring

upon unfolding. To that end, we employ a protein construct that we have designed and characterized recently, see Fig. 3.1.[15,184–186] That is, we have chosen the second PDZ (PDZ2) domain from human tyrosine-phosphatase 1E (hPTP1E), which has been studied extensively as a model allosteric protein from different perspectives, i.e., structural,[114,187,188] dynamical[112,189,190] or computational.[15,29,30,75,128,186,191–196] In order to investigate the allosteric mechanism by transient IR spectroscopy, we have covalently linked an azobenzene derivative across the binding groove of the PDZ2 domain in such a way that the light-driven *trans-cis* isomerization of the photoswitch induces a structural transition in the protein, which mimics ligand unbinding in the native system.[184] By NMR analysis (PDB entries: 2M0Z and 2M10), we have confirmed that the structural changes, which are of the order of 1 Å, are of similar size as in the native system.

We have shown by transient IR spectroscopy that the protein responds to the light-induced perturbation on two timescales, one extending up to ≈100 ns that reflects the opening of the binding groove (which we know since luckily we could isolate one specific mode localized on the azobenzene derivative), and a second phase on a 10 μs timescale.[184] We have no information from experiment on the structural nature of the second phase, since we did not employ any label that would reveal site-selective information. Based on molecular dynamics (MD) simulations, we suggested that it involves some of the more flexible and remote loop regions of the protein and/or the termini.[15,186] MD simulations also suggested[184] that the water solvation shell changes on the timescale of the binding groove opening even relatively far away from the perturbation, and we proposed that both aspects (i.e., structural changes of regions far away from the binding groove and/or changes in solvation) might be possible mechanisms of allostery. Based on the experience from previous works employing AHA in other molecular systems,[70,71,93,94,176–178] it appears possible that this label can sense such effects in a photoswitchable PDZ2 domain, and it is the goal of the present paper to explore whether this is indeed possible.

As a first step in this direction, we consider stationary difference spectra comparing the two states of the protein, and leave time dependent transient experiments for a future publication. To explore the capability of the AHA label as a probe of the change of the local structure, the following criteria have been used to select the positions for the AHA label:

- Amino acids close (L78, N80, I20) *versus* far away (N16, S48, L66) from the photoswitch have been selected, in order to investigate the dependence on the distance from the perturbation.

- Amino acids within secondary structural elements (L78, S48) *versus* loop regions (N16, L66) have been selected. Whereas the PDZ2 domain undergoes only small shifts in the more rigid secondary structural elements, larger conformational changes can be observed in the flexible loops.[184]

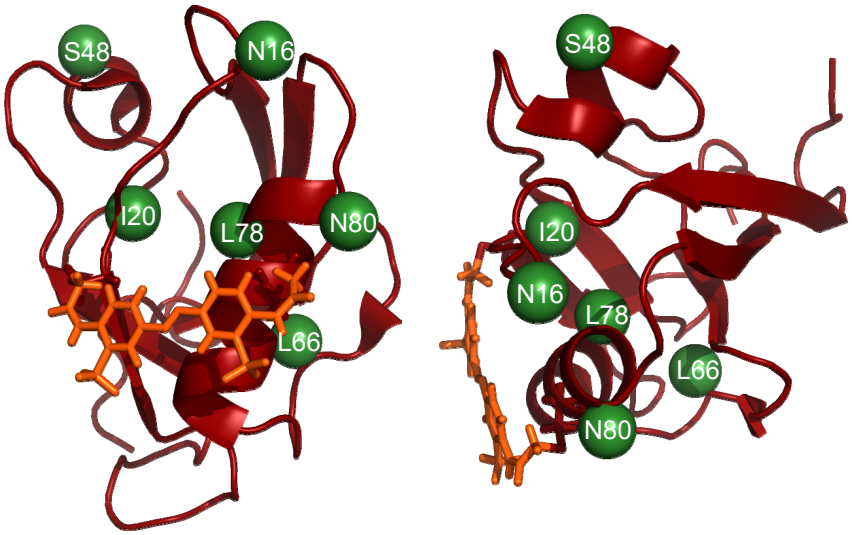


FIGURE 3.1: Photoswitchable PDZ2 domain with the positions of amino acids, which have been mutated to AHA, indicated in green. Only one amino acid has been mutated at a time in the experiment. Left: Front view to the binding site of the protein; right: top view

- Surface exposed amino acids (N80, N16, S48) *versus* amino acids buried in the hydrophobic core (L66, I20) have been selected. Only the former will sense changes in protein solvation.
- Finally, we considered only amino acids of similar size for mutation, and only neutral amino acids in order to avoid shifts of the isoelectric point of the protein.

Fig. 3.1a,b show all amino acids that have been mutated to AHA (for each mutation, only one amino acid at a time). We use 2D IR spectroscopy instead of FTIR spectroscopy owing to the inherent sensitivity gain of 2D IR, which for the most part originates from its quadratic dependence on the extinction coefficient that significantly reduces the solvent background in a relative sense.[91] Furthermore, since the measurement beams are small, and since we can light-induce the *trans*-to-*cis* transition, extraordinary low amounts of protein sample are needed ($\lesssim$ 1 nmol) in 2D IR difference spectroscopy. These low amounts compensate for the fact that protein preparation is quite tedious.

3.2 Material and Methods

3.2.1 AHA-Labelled, Photoswitchable PDZ2 Domain

To produce protein containing the non-natural amino acid AHA, we have modifed a protocol that has been described before[183] as follows: Starting from the previously used pET30a(+) vector containing two mutations S21C and E76C for the cross-linking of the photoswitch,[184] additional single residues were mutated to methionine for the insertion

of AHA by site-directed mutagenesis (QuikChange; Agilent, Santa Clara, CA). A methionine auxotrophic cell strain (E. coli B834(DE3); Novagen, Merck Millipore, Darmstadt, Germany) was used for the protein expression. Cell cultures were grown in LB medium with 30 μg/ml kanamycin at 37°C to an OD600 of 0.8. The LB cultures were centrifuged at 3500 g and 20°C for 20 min and afterwards immediately re-suspended in a minimal medium (SelenoMet Base medium with SelenoMet Nutrient mix from Molecular Dimensions, Newmarket, UK) supplemented with 100 μg/ml AHA (Bapeks, Riga, LV) and 30 μg/ml kanamycin, where they were incubated at 37°C for 30 min in order to use up residual methionine. 1 mM IPTG was added and the protein was expressed at 37°C. AHA is a potentially reactive amino acid and can be modified during protein expression and purifcation. A short expression time of 4 hours was therefore chosen in order to minimize protein modifications, despite a somewhat lower yield. The cell cultures were centrifuged at 3500 g and 4°C for 20 min and then stored at -20°C.

The His-tagged protein was purified using nickel magnetic beads (Biotool.com, Houston, TX, USA) following the recommended protocol. 50 mM Tris buffer, pH 8.5 was used for cell lysis, the protein was denatured with 6 M guanidinium chloride and then refolded on the nickel magnetic beads. Residual nickel, which was bound to the protein, was removed by adding 40 mM EDTA and incubating over night at 4°C. A yield of about 7 mg protein per liter of cell culture was determined using a Bradford assay. SDS-PAGE and mass spectrometry were performed to control the purity of the protein.

As photoswitch, we employed an azobenzene derivative containing two sulfonate (-SO_3^-) groups in order to increase its solubility in water.[197] For its cross-linking, a protocol similar to the one described in Ref.[184] was used. However, the reduction of the cysteins, to which the photoswitch binds, had to be performed under milder conditions, since any reducing agent that is commonly used for that purpose would also reduce the azido group of the AHA, resulting in a primary amine. The reaction conditions had to optimized, so that most of the disulfide bridges were reduced while most of the AHA remained intact. That is, 1 mM TCEP (from a 100 mM TCEP stock at pH 8.5) was added to the protein at a concentration of around 50 to 200 μM in 50 mM Tris buffer, pH 8.5, 500 mM imidazole and 40 mM EDTA. This solution was incubated at room temperature for no longer than 15 min. In principle, the reduction of disulfide bonds could be omitted completely in order to avoid any destruction of AHA, which however, would lower the yield of the cross-linking reaction dramatically.

As in Ref.,[184] the reducing agent was subsequently removed by desalting chromatography (HiPrep column, GE Healthcare, 50 mM Tris buffer, pH 8.5), and the protein was cross-linked to the photoswitch under an oxygen free (nitrogen) atmosphere at room temperature for at least 6 hours. Cross-linking was performed in a highly diluted solution in order to minimize the formation of oligomers (10 μM protein to 100 μM photoswitch). The cross-linked protein was purified using anion exchange chromatography (HiTrapQ column, GE Healthcare). The His-tag was removed by digestion with

HRV 3C protease, and the cleaved protein was purifed with nickel affinity chromatography (His-Trap HP column, GE Healthcare). The purified protein was concentrated and desalted into 50 mM borate buffer with 150 mM NaCl at pH 8.5. Finally, the protein was lyophilized and dissolved in D_2O. Mass spectra of all mutants considered in this study are shown in Fig. S1 (Supplementary Material), emphasizing the excellent purity of the final labelled and cross-linked protein.

3.2.2 Difference 2D IR Spectroscopy

For the difference 2D IR spectroscopy, we used an instrument described before.[93,94] In brief, mid IR pulses were generated in a home-built two stage OPA with a difference mixing stage[198] pumped by a commercial Ti:S amplified laser system (Spitfire, Spectra Physics) running at 5 kHz. The OPA yielded pulses at 4.7 μm and $\approx$3 μJ per pulse with an energy stability better than 0.3% at 500 shots. The 2D-IR instrument used a four wave mixing phase-matching geometry employing a HeNe trace beam to accurately determine the delay times[199] and a polarization-based balanced heterodyne detection.[200] The signal was detected on a 2$\times$32 MCT detector array after dispersing it in a spectrograph with a resolution of 7 cm^{-1}. A photoelastic modulator (PEM) was used to induce a quasi-phase shift on pulses 1 and 2 in order to suppress scattering.[201] The time domain data were collected into 2.11 fs long time bins (defined by the HeNe wavelength) with a maximum scanning time of 3 ps, revealing a spectral resolution of 2.7 cm^{-1} after zero-padding by a factor 2 and subsequent Fourier transformation. Purely absorptive spectra were obtained by alternative scanning of beams 1 and 2 backward in time. The population time was kept constant at 300 fs to minimize non-resonant effects from overlapping excitation pulses.

At the protein concentrations considered in this study (1 mM or below), the AHA signal is buried under the background from the D_2O buffer (essentially the wing of the OD-stretch vibration that is centered at $\sim$2500 cm^{-1}). We therefore measured difference 2D IR spectra, i.e., we first measured a spectrum of the dark adapted protein, in which case the photoswitch is in the *trans* configuration. The sample was then switched into *cis* by illuminating it at 370 nm from a cw-diode laser (CrystaLaser CL-2000) for about 3 min, and a second 2D IR spectrum was measured without changing any of the alignment (but with continuous illumination with the 370 nm laser). The difference of the two spectra was then calculated after phasing them independently, using the phase of the water background also contained in the data as a reference (see Ref.[202] for details). Since the water background is the same in that light-induced difference spectrum, it did not have to be measured independently and a stationary cuvette with only $\approx$ 1 μl of sample volume was sufficient for these experiments. For the unfolding difference spectrum, on the other hand, L78AHA was resuspended in 6 M guanidinium chloride, deuterated buffer solution at pH 8.5 to ensure complete unfolding. The final spectrum is a double-difference spectrum, i.e. the spectrum in the unfolded state minus that of the

corresponding buffer, subtracted from the folded state minus corresponding buffer. To subtract out the buffer contribution, a syringe pump sample delivery system together with a flow cell was used to exchange sample,[93,94] requiring much larger sample volumes of $\approx 100\ \mu$l.

3.2.3 Molecular Dynamics Simulations

Recently Buchenberg et al. performed a detailed molecular dynamics (MD) study of the structural changes of photoswitchable PDZ2 upon *cis-trans* photoisomerization.[186] Following this work, we carried out MD simulations of photoswitchable PDZ2 including AHA labels as present in the experiment. To minimize computational time, three labels were considered per simulated system (in contrast to experiment, where only one amino acid was replaced per sample). Selecting label positions such that no AHA group interacts with another one, the first system contained AHA labels at sequence positions 16, 66 and 78, and the second system at positions 20, 48 and 80. Both systems were simulated in the *cis* and *trans* configuration of PDZ2, using Gromacs 4.6.7 with a hybrid GPU-CPU acceleration scheme.[203]

In all simulations, the protein was placed in aqueous solution including 150 mM NaCl. The side chains of all four histidine residues (33, 54, 72, and 87) in all initial structures were chosen to be ϵ-protonated. The protein was described using the Amber99SB*ILDN force field,[139–141] water molecules by the TIP3P model,[148] and ions with the model of Ref..[204] The parameterization procedure using Antechamber[205] and Gaussian09[143] and the resulting force field parameters of the AHA labels are described in the Supplementary Materials. All bonds involving hydrogen atoms were constrained,[206] allowing for a 2 fs time step. Electrostatic interactions were calculated using PME.[145] The minimum cutoff distance for electrostatic and van der Waals interactions was set to 1.2 nm. To couple the system to a heat bath, we used the velocity-rescale algorithm[146] and for pressure coupling the Berendsen algorithm.[147] After energy minimization, the systems were simulated for 100 ns at a pressure of 1 bar and a temperature of 300 K.

Data evaluation was carried out with Gromacs tools.[203] To determine intraprotein polar contacts, we used *g_mindist* to calculate the minimal distances between the $-N_3$ atoms of the AHA residue and all polar atoms which were found within a 1 nm radius of the azido group (with respect to the starting structure). Contact distributions were then obtained by histogramming the MD data with 0.01 nm binning width. We defined a contact to be formed if the minimal distance between a azido group nitrogen atom and a protein nitrogen or oxygen atom is shorter than 0.45 nm.[52] In a similar way, we analyzed contacts between AHA and water as azido group/water oxygen atom distances with a cut-off of 0.45 nm, as well.

3.2.4 Calculation of Vibrational Spectra

We used the empirical model of Cho and coworkers[70] to estimate vibrational shifts $\delta\omega$ caused by changes in the electrostatic environment of the AHA labels. By calculating the electric field $\mathbf{E}_j(t)$ at the nitrogen atoms ($j = 1, 2, 3$) of the azido group for each MD snapshot at time t, we obtain the spectral shift

$$\delta\omega(t) = \sum_j \mathbf{a}_j \mathbf{E}_j(t) \tag{3.1}$$

with coefficients $\mathbf{a}_j$ given in Ref..[70] Electric fields were computed via a reaction field approach[203,207] using a cut-off radius $r_c = 2.3$ nm. From the frequency trajectory $\delta\omega(t)$ with a time step of 15 ps, the distribution of the vibrational shifts was obtained via a histogram using 50 bins.

As an alternative approach, quantum-mechanical calculations of vibrational spectra of protein side-chain conformers were performed using Gaussian09,[143] following Wolf et al.[208] From the MD simulation, we first determined the snapshot where the $N^{(2)}$-atom of the azido group is closest to one of the two sulfur atoms of the photoswitch sulfonate groups; we chose one structure each for the *trans* and the *cis* configuration, and for I20AHA and L78AHA. Using these structures, we constructed a minimal vacuum model of the AHA label ($-N_3$ group and Cβ atom) and the sulfonate group (including the attached carbon atom), where both carbon atoms were saturated with hydrogen atoms. Keeping the positions of both carbon atoms as well as the distance between $N^{(2)}$-atom and sulfur atoms fixed, the resulting system was initially minimized in energy at the HF/6-31+G* level, followed by a density functional theory-based minimization using B3LYP[209,210] and the 6-31+G* basis set. At the same theoretical level, harmonic frequencies and band intensities were determined by diagonalization of the Hessian matrix. Vibrational frequencies were corrected by the asymmetry factors given in Ref..[211] Atomic charges were calculated via Mulliken population analysis.[212]

3.3 Results and Discussion

3.3.1 Folding Stability

We have determined the folding stability of the various mutants by CD spectroscopy; the data are shown in Fig. S2 (Supplementary Material) and the results are summarized in Table 3.1. The CD measurements have been performed under the same conditions as the 2D IR experiments, i.e., buffered in D_2O solution. Tentatively, the AHA mutations destabilize the protein, if at all, only a little bit relative to the photoswichtable PDZ2 domain without AHA, as judged from the midpoint temperatures T_m. The *cis*-state is, overall speaking, somewhat less stable than the *trans*-state (the latter is opposite to what we have reported in Ref.,[185] where the CD has however been measured in H_2O and

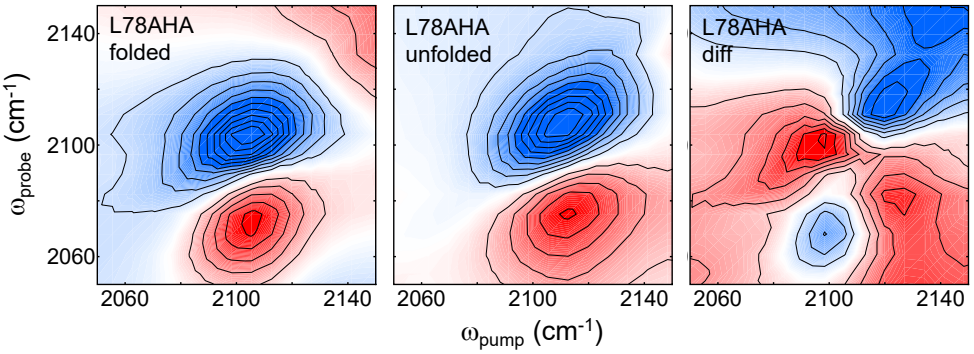


FIGURE 3.2: Purely absorptive 2D IR spectra of L78AHA in the folded (left panel) and the unfolded state (middle panel), the latter induced by adding 6 M guanidinium chloride. The right panel shows the difference (unfolded minus folded) of both spectra. To that end, the spectra of the folded and unfolded protein have been scaled to the bleach signal, since they have been measured at slightly different concentrations, and have subsequently been subtracted to reveal the difference spectrum. For the plotting, the difference spectrum has been multiplied by a factor 2 so that all spectra share the same number of contour lines. Blue colors depict negative signals (i.e., bleach and stimulated emission in the purely absorptive 2D IR spectra) and red colors positive signals (excited state absorption).

without NaCl). Furthermore, the folding transition is less cooperative in the *cis*-state with a larger width of the folding transitions ΔT. Nevertheless, all mutants considered here are folded to $\gtrsim$98% at 10°C in both their *cis* and their *trans*-states. We performed the 2D IR experiments at that temperature in order to ensure that any difference signal induced by photoswitching is not obscured by partial unfolding.

TABLE 3.1: Unfolding midpoint temperatures T_m and width of the folding transitions ΔT of the different mutations of the photoswitchable PDZ2 domain in its two states, as obtained from CD spectroscopy. To that end, the data have been fit to a function $1/(1+\exp((T-T_m)/\Delta T))$ after subtraction of the background and normalisation.

Mutant	***trans***		***cis***	
	T_m/°C	ΔT/°C	T_m/°C	ΔT/°C
no AHA	49.5±1	3.7±1	46.0±2	6.3±2
N16AHA	48.0±1	5.4±1	45.5±2	8.0±2
I20AHA	43.0±1	5.2±1	36.5±4	7.9±2
S48AHA	49.5±1	3.9±1	51.0±1	3.8±1
L66AHA	44.5±1	4.2±1	43.5±1	5.9±1
L78AHA	43.5±1	4.4±1	38.5±2	7.4±2
N80AHA	47.0±1	4.5±1	49.0±3	6.0±3

3.3.2 Unfolding 2D IR Difference Spectra

As a reference experiment, Fig. 3.2 compares the 2D IR spectra of L78AHA in the folded and the unfolded state of the protein. The measurements were performed in the dark, i.e., with the photoswitch in its *trans* configuration. We chose to induce unfolding by a denaturant (6 M guanidinium chloride), rather than by raising the temperature, as the latter causes a dramatic change in the water response and furthermore tends to induce aggregation of the protein and hence scattering in the 2D IR signal. In each case, a 2D IR spectrum of the corresponding buffer has been measured as well under identical conditions, and has been subtracted, as shown in Refs..[91,93,94] In both states of the protein, the 2D IR spectra show the usual 0-1 peak depicted in blue together with the 1-2 peak depicted in red (i.e., with opposite sign), which is shifted along the probe-frequency axis due to the anharmonicity of the AHA vibration. By the tilt of the 2D IR lineshapes, a modest amount of inhomogeneity is detected, which does not differ very much in the two states of the protein.

The most prominent change upon unfolding is a blue-shift of the AHA label by $\approx$7 cm^{-1}, which is a bit smaller than for previous observations.[176] That is, upon unfolding, the AHA label becomes fully solvated and hence the number of hydrogen bonds to water molecules increases, as well as their flexibility allowing for hydrogen bonding at a more optimal angle. Cho and coworkers have shown in Ref.[71] that both effects cause a blue shift of the vibrational transitions. In turn, the frequency shift also reveals that the AHA label of this particular mutant is solvent-exposed to only a minor extent in the folded state of the protein. Position L78AHA is situated inside the the binding pocket (see Fig. 3.1), where the access of the solvent is limited, possibly shielded by the azobenzene photoswitch (see below).

3.3.3 Photoswitching 2D IR Difference Spectra

With that information in mind, we turn to the 2D IR difference spectroscopy induced by photoswitching the azobenzene moiety from *trans* to *cis* with the help of a cw-laser diode. Fig. 3.3 shows the spectra of all mutants that have been investigated. Surprisingly, only one of the considered mutants (L78AHA) reveals an evaluable signal, while all others show no clear signal apart from some modifications of the water background and/or small remaining scattering (the latter appears as spurious signals along the diagonal, see e.g. the 2D IR difference spectra of L66AHA or N80AHA). These two effects currently limit our sensitivity, not signal-to-noise *per se*. It should however be stressed that we know beyond any doubt that the AHA label is present also in those mutants, for which no difference signal could be detected. For example, we see its absorption band in the individual dark-adapted (*trans*) 2D IR spectra, i.e., before taking the difference with the corresponding *cis* 2D IR spectra (however, sitting on a large water background,

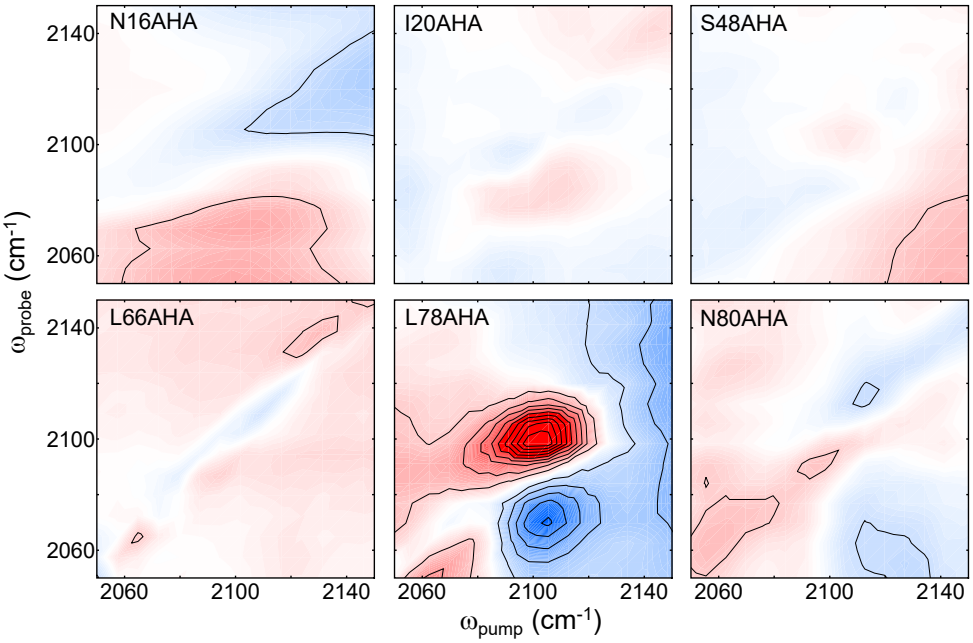


FIGURE 3.3: Purely absorptive 2D IR difference spectra (*cis* minus *trans*) of all mutants considered in this study, induced by photoswitching the azobenzene moiety from *trans* to *cis* with the help of a cw-laser diode. All signals were normalized by concentration (which varied between 0.6 mM and 1.1 mM) and to the peak signal of L78AHA. Blue colors depict negative signals and red colors positive signals.

see Fig. S3, Supplementary Material). Also mass spectrometry (Fig. S1, Supplementary Material) confirms the existence of an azido group in all mutants.

Nonetheless, the response of L78AHA, and the comparison to the unfolding difference spectrum from Fig. 3.2, is quite revealing. The 2D IR difference spectrum induced by photoswitching (Fig. 3.3) is quite comparable to the 2D IR spectrum of the folded state (Fig. 3.2, left panel), where the frequency position and 2D line shape is concerned, but the sign (encoded by the colors) is inverted and the intensity is about a factor 5 smaller. Furthermore, the photoswitch induced difference spectrum is very different from the unfolding induced difference spectrum (Fig. 3.2, right panel), the latter of which resulting from the frequency shift of the AHA transition. That is, the effect of photoswitching on the AHA label is mostly a reduction of the vibrational transition dipole in the *cis*-state without affecting the vibrational frequency very much. This in turn also evidences that the difference spectrum induced by photoswitching is not the result of the slightly reduced stability of the protein with the photoswitch in the *cis* state, which according to CD spectroscopy (Supplementary Material, Fig. S2) might cause ≈2% of unfolding.

To the best of our knowledge, this effect has not been described so far. That is, while the frequency of the AHA label is considered to be a measure of the amount of solvation,[71,93,94,175–178] the intensity stays essentially the same, which is indeed what

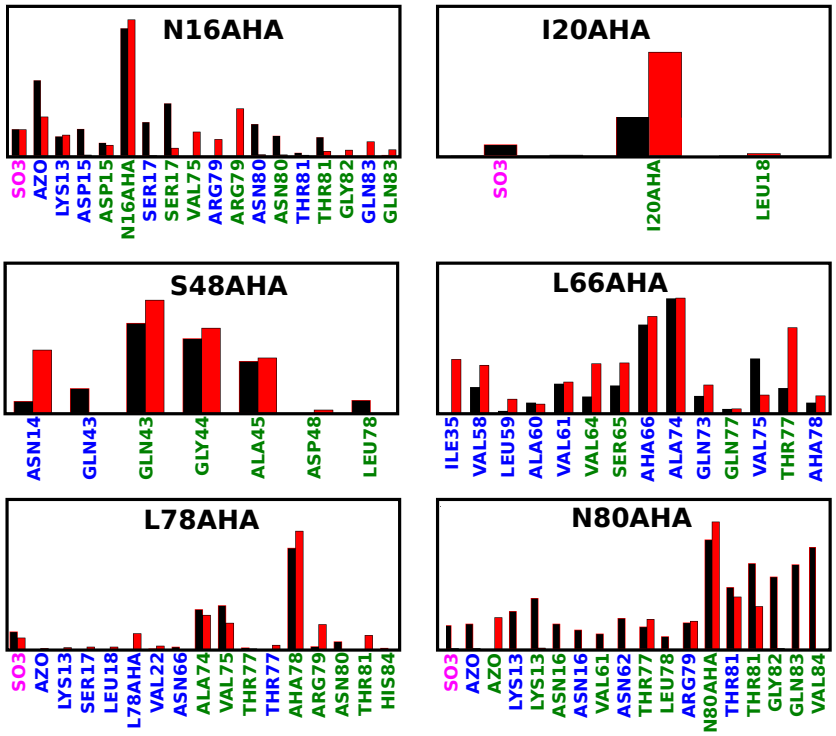


FIGURE 3.4: Histograms of polar contacts between considered AHA labels and PDZ2 protein in *cis* (black) and *trans* (red) state. Green and blue labels denote backbone and side-chain contacts, respectively, and magenta labels indicates contacts with the sulfonate groups of the photoswitch.

is observed when unfolding the protein (Fig. 3.2). It is conceivable that when the label enters a more heterogenous environment, the absorption becomes wider at the expense of the peak intensity such that the integrated intensity, and hence the transition dipole, stays the same. We tentatively exclude that effect here, since it would cause wings of opposite sign on both sides of the peak in the difference spectrum that are not observed, even when considering the present signal-to-noise level. We also carefully checked the possibility that the loss of intensity reflects a loss of AHA label due to a reduction of the azido group, for example. To that end, we first measured mass spectra before and after laser illumination (see Fig. S4, Supplementary Material), showing that no chemical modification is occurring, such as the loss of a N_2 molecule. Second, we measured FTIR difference spectra switching forth and back, evidencing that the transition is indeed reversible and that the AHA band regains its intensity upon *cis-trans* back-switching (see Fig. S5, Supplementary Material).

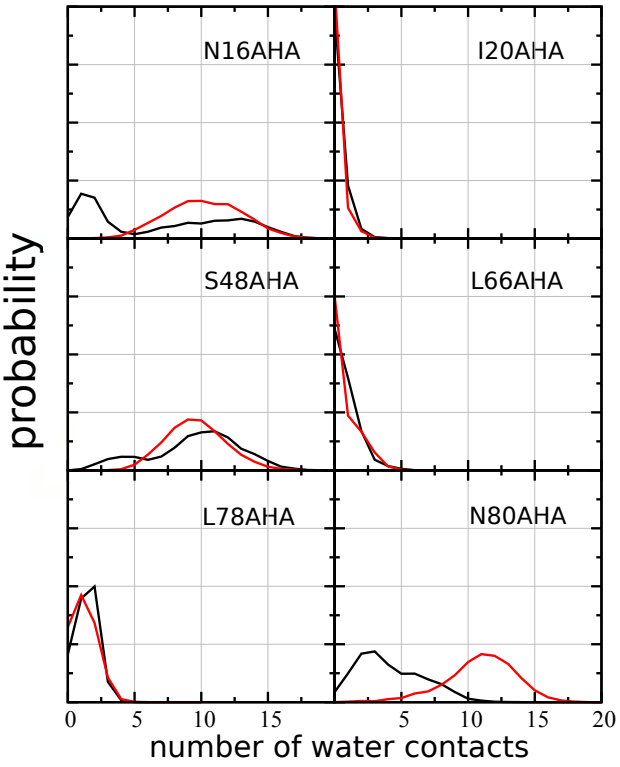


FIGURE 3.5: Distribution of the number of water contacts of the AHA labels, with the protein being in *trans* (red) and *cis* (black) state.

3.3.4 Spectral Simulations

To get insights into the molecular mechanism giving rise to this spectroscopic response, we performed spectral simulations based on all-atom MD simulations. Recent quantum-chemical calculations of Cho and coworkers[71] have shown that the spectroscopic signatures of the azido stretch mode of AHA mainly reflect the local electrostatic environment of the azido group. To result in an observable IR difference spectrum of PDZ2, this electrostatic field needs to differ in the *trans* and the *cis* states. Since the main contributions to the local electrostatic environment arise from contacts with nearby polar residues and water molecules, we first employ MD simulations to study possible changes of these contacts caused by the *trans-cis* isomerization.

Let us begin with the contacts between polar protein residues and the azido groups. Fig. 3.4 reveals that the number of such intraprotein contacts and their changes upon *trans-cis* photoisomerization appear quite complex. Roughly speaking, we find three labels (N16AHA, L66AHA, N80AHA) with a large number, two (S48AHA, L78AHA) with a medium number, and one (I20AHA) with a small number of intraprotein contacts. Notably, we see that labels N16AHA, I20AHA and N80AHA show significant contact changes between *trans* and *cis* states, while the label L78AHA shows only minor

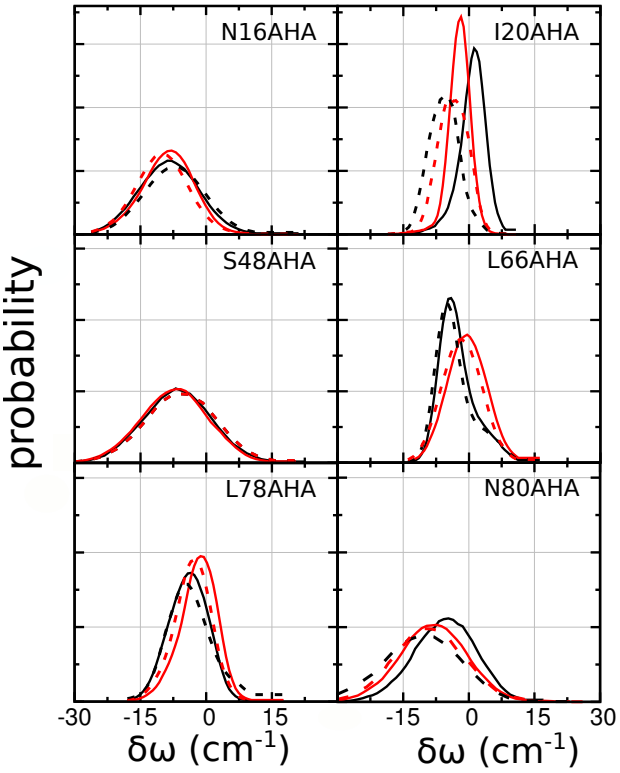


FIGURE 3.6: Vibrational frequency shifts (relative to the gas-phase vibration) of the azido stretch mode of the AHA labels in the *trans* (red) and *cis* (black) states. Results are obtained from the electrostatic model (Eq. (3.1)) with (full lines) and without (dashed lines) inclusion of the sulfonate groups of the photoswitch.

variations. Fig. 3.5 displays the distribution of water contacts during the MD simulation. Three AHA labels (N16AHA, S48AHA, N80AHA) are found to be strongly hydrated, two (I20AHA, L66AHA) are hardly hydrated, and label L78AHA, for which a difference signal could be observed experimentally, may be characterized as moderately hydrated. Interestingly, only the strongly hydrated labels show significant differences in the number of water contacts between *trans* and *cis* states. In particular, we find that the decrease in the number of water contacts of N16AHA and N80AHA upon *trans-cis* photoisomerization is compensated by an increase of protein contacts.

As the examination of the changes of contacts does not provide a clear explanation of the experimental findings, we next study to what extent the contact changes discussed above are reflected in spectral changes of the azido stretch mode of the corresponding AHA label. To this end, Fig. 3.6 displays the distribution of vibrational frequency shifts $\delta\omega(t)$ upon *trans-cis* photoisomerization, as obtained from the electrostatic model in Eq. (3.1). In the case of the strongly hydrated labels (N16AHA, S48AHA, N80AHA), the calculations predict relatively broad ($\sim \pm 10\,\mathrm{cm}^{-1}$) and red-shifted (by $\sim 10\,\mathrm{cm}^{-1}$) frequency distributions. The weakly to hardly hydrated labels (I20AHA, L66AHA, L78AHA), on the other hand, exhibit a smaller distribution width ($\sim \pm 5\,\mathrm{cm}^{-1}$) and

also smaller frequency shifts. This different width of the frequency distributions of hydrated vs. not hydrated labels is in nice agreement with the experimental results of Taskent-Sezgin et al.,[176] who found line widths of $\sim \pm 5$ and $10\,\mathrm{cm}^{-1}$ for AHA labels in the folded and unfolded (i.e., water exposed) state of a protein, respectively. On the other hand, we note that the overall redshift predicted by the model is in variance with the common expectation that solvation rather causes a blue shift (see, e.g., Fig. 3.2 as well as Refs.[88,93,94,176,178]). As discussed in Ref.[70] [Fig. 6], this effect is caused by our neglect of the polarizability of the azido group.

Upon *trans-cis* photoisomerization, the strongly hydrated labels show no or only minor spectral changes. This is clearly expected for S48AHA, where both protein and water contacts remain more or less unchanged, but comes a bit as a surprise for N16AHA and N80AHA, where both protein and water contacts change significantly. We conclude that the highly mobile water molecules around these labels may effectively screen or counter the electrostatic interactions of the protein, which hampers a clear spectroscopic response. The frequency distribution of L78AHA hardly changes between *trans* and *cis* states, which is in line with the absence of major protein or water contact changes for this label. On the other hand, the spectral simulations would predict a significant blue shift of I20AHA and L66AHA when changing from *trans* to *cis*, resulting from the changes of protein contacts, which however is not observed experimentally.

To sum up the results up to this point, our MD simulations combined with the electrostatic model by Cho and coworkers[70] would predict spectral changes for I20AHA and L66AHA, while in experiment L78AHA is the only label with an observable spectral response upon *trans-cis* photoisomerization. Of course, one may question the accuracy of the structural prediction of the employed MD force field (neglecting, e.g., the polarizability of the azido group) or the assumptions underlying the electrostatic model (neglecting, e.g., the dependence on dispersive interactions or the fact that the model of Cho and coworkers[70] has been parameterized based on QM calculations with water clusters around an azido group only, while intraprotein contact to polar residue play a significant role here as well). Moreover, the electrostatic model would describe frequency shifts only, which is not what is observed experimentally. That is, the experimental 2D IR difference spectrum upon *trans-cis* photoisomerization is dominated by a change in the intensity of the vibration of L78AHA, and not its frequency.

Being a 1,3-dipole, the azido group is strongly polarizable, and it is for example well known that azido groups in different molecules have strongly varying transition dipoles.[92] Hence, it is conceivable that even weak electronic interactions with other parts of the protein, involving for example a charge transfer similar to the situation in a hydrogen bond or at the onset of a nucleophilic attack, may strongly affect the electronic density of the azido group and thereby change the transition dipole of its vibration. The prime candidates for such an electronic interaction are the two sulphonate groups of the photoswitch, which have been introduced to increase the solubility of the azobenzene

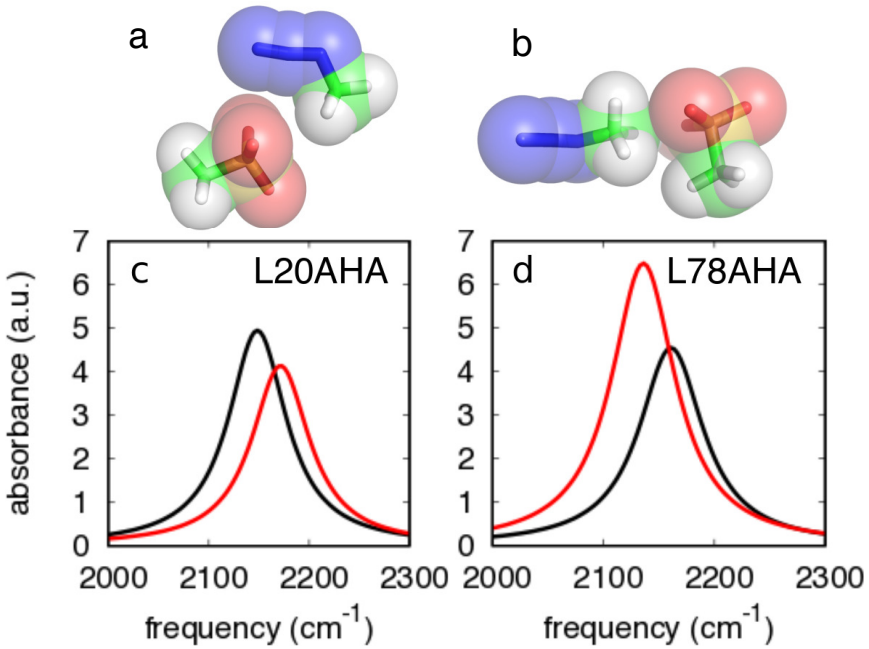


FIGURE 3.7: Quantum-chemical calculations for a minimal model of the AHA label ($-N_3$ group and Cβ atom) and the sulfonate group of the photoswitch, indicating the van der Waals radii of the atoms by colored spheres. Panels (a) and (b) show energy-minimized structures of L78AHA in *cis* with contact and *trans* without contact, respectively, of the azido and the sulphonate groups. I20AHA exhibited similar structures, with a contact in *trans* and without contact in *cis* (not shown). Using these structures, vibrational spectra of the azido stretch mode were calculated for (c) I20AHA and (d) L78AHA. Black and red lines correspond to the *cis* and *trans* state of the photoswitch, respectively.

moiety.[197] Performing electrostatic calculations in the absence of these sulfonate groups (dashed lines in Fig. 3.6), the frequency distributions of I20AHA and L78AHA reveal a significant red-shift, emphasizing the importance of the sulfonate groups. Moreover, the red-shifts are clearly different for *trans* and *cis* states. The strongly hydrated labels N16AHA and N80AHA, on the other hand, show only minor spectral changes when sulfonate groups are excluded in the electrostatic calculations. Although these labels form contacts with the sulfonate groups (Fig. 3.4), the strong electrostatic screening due to solvent water appears to prevent clear spectral changes.

In our MD simulations, such contacts don't appear often (cf. Fig. 3.4), owing to the missing stabilizing electronic interaction in the force field description. To investigate if such direct contacts between azido and sulfonate groups can lead to a charge transfer and the experimentally observed absorbance changes, we extracted the structure with the closest distance between I20AHA/L78AHA and an neighboring sulfonate group from MD simulations of each *cis* and *trans* states. Showing the energy-minimized structures of these states, Figs. 3.7a and b reveal that L78AHA indeed forms a close contact

with the sulfonate group, which mediates an electronic interaction. Intriguingly, this contact only exists in the *cis* configuration of the photoswitch, but not in the *trans* configuration. Likewise, an energy-minimized structure of I20AHA (not shown) exhibits a close contact associated with an electronic interaction in the *trans* configuration, that does not exists in the *cis* configuration.

Using these structures, we calculated harmonic frequencies and band intensities of the azido stretch mode (see Methods). Figure 3.7c and d show the resulting vibrational spectra of the azido stretch mode of I20AHA and L78AHA in the *cis* and *trans* states of the photoswitch. In either case, a close contact between the azido and the sulfonate groups results in a decrease of the band intensity of the AHA vibration due to a charge transfer to the azido group that reduces the transition dipole. As can be seen in Table 3.2, the charge on all nitrogen atoms changes by up to $\pm 0.23\,e$, indicating that the charges on the azido group are indeed very mobile due to electronic effects. A polarizable force field is the least that would be needed to describe that effect self-consistently in a MD simulation. The effect on intensity is stronger for L78AHA than for I20AHA, in qualitative agreement with experiment. It should however be mentioned that in addition to the intensity change, the quantum chemistry calculations would also predict a frequency shift of ~ 25 cm^{-1}, which is not observed in experiment. Possibly, the missing protein environment and solvation in the quantum chemistry calculations might be responsible for that effect. The results nevertheless suggest that weak electronic interactions may indeed change the intensity of the AHA vibration, enabling a new application for that vibrational label.

3.4 Conclusion

The AHA vibration is an attractive label to study protein structure and dynamics with the help of IR and 2D IR spectroscopy, since it can be incorporated almost anywhere in a protein by a methionine auxotrophic protein expression strategy,[182,183] since its transition dipole is reasonably large for 2D IR measurements in the sub-mM regime,[91,94] and

TABLE 3.2: Mulliken charges[212] (in units of e) of the AHA azido nitrogen atoms in different quantum chemical vacuum models. Cont./no cont. refer to the presence/absence of a contact with the neighboring sulfonate group. $N^{(3)}$: nitrogen attached to AHA side chain; $N^{(2)}$: internal nitrogen atom; $N^{(1)}$: terminal nitrogen atom.

nitrogen atom	**I20AHA**		**L78AHA**	
	cis **no cont.**	***trans*** **cont.**	***cis*** **cont.**	***trans*** **no cont.**
$N^{(3)}$	-0.45	-0.33	-0.52	-0.50
$N^{(2)}$	1.02	1.21	1.08	0.89
$N^{(1)}$	-0.71	-0.94	-0.67	-0.54

since it is a sensitive probe of its solvation environment.[70,71,88,93,94,176–178] However, in order to demonstrate the last aspect, rather dramatic changes have been applied to the label so far. For example, a protein has been unfolded, which fully exposes an AHA label that normally is situated in the hydrophobic core of the protein to water,[176] or a peptide ligand with an AHA label has been dissociated from a protein,[93,94] which has about the same impact on the label as unfolding. Employing a combined experimental-computational approach, we set out in the present work to explore the capability of the AHA label to also sense much smaller changes in the structure and solvation of a mid-sized protein. That is, rather than unfolding of a protein, we change the structure of a PDZ2 domain in a very modest way ($\lesssim$ 1 Å) with the help of a photoswitch that is covalently linked to it, thereby mimicking the conformational transition upon ligand binding.[15,184,186]

To that end, we first had to develop a protocol for the post-translational synthesis of the cross-linked PDZ2 domain containing AHA labels. That is, both the cysteins needed for the linking of the azobenzene photoswitch as well as the azido groups of AHA are chemically very reactive groups. For AHA, this reactivity has been used successfully for click chemistry reactions,[213] however, any modification of this unnatural amino acid had to be avoided in our case. Mainly, the reduction of disulfide bridges had to be optimized, such that AHA would not be reduced at the same time into a primary amine.

We have explored six mutations, distributing the AHA label at various positions of the protein that differ significantly in their properties (Fig. 3.1), e.g., inside the hydrophobic core *versus* surface exposed, or in the flexible loops *versus* in more rigid secondary structure motives. We find that because of the fact that the label is very small and of medium polarity, it can replace both polar and apolar amino acids without affecting the stability of the protein too much (Table 3.1). In that sense, it is indeed a versatile label.

Somewhat surprisingly, however, we observe an evaluable difference 2D IR signal upon photoswitching only for one mutation, L78AHA (we nevertheless chose to show all results in Fig. 3.3, including the negative ones, since the purpose of this survey has been to learn what can be sensed with the AHA label, and what its limitations are). The very distinct difference signal of only the one label L78AHA comes as a surprise, since I20AHA and N80AHA are equally close to perturbation introduced by the photoswitch. Naturally, one would assume that the size of the effect correlates with the distance to the perturbation; yet, with our current experimental sensitivity we cannot detect any response for I20AHA and N80AHA. It also comes as a surprise, since the MD simulations would in fact predict significant changes of local contacts to both water molecules and protein side chains, even for labels that are quite far away from the photoswitch, such as N16AHA or L66AHA (see Fig. 3.4 and 3.5).

The distinct response of L78AHA, in turn, suggests that it results from a very specific interaction that should be robustly reproduced by a MD simulation, even given the

unavoidable imperfectness of any MD force field that might describe some structural details in not quite the correct way. We therefore first employed classical MD simulations, calculating local contacts of the various AHA labels with its surrounding (Figs. 3.4 and 3.5) as well as electric field induced frequency shifts along the line of a model put forward by Cho and coworkers (Fig. 3.6).[70] The simulation revealed that the strongly hydrated labels show no or only minor spectral changes, even if both protein and water contacts change significantly (as is the case for N16AHA and N80AHA). This suggests that this is the result of a mutual cancellation of the contributions from water and the protein. That is, the highly mobile water molecules around these labels effectively screen or counter the electrostatic interactions of the protein.

On the other hand, the calculations revealed nothing with respect to which L78AHA would stick out. In the contrary, they predict a frequency shift for I20AHA and L66AHA that are of the same order as in the unfolding experiment of Fig. 3.2 or what has been observed in Ref.[94] for ligand unbinding, and thus should be measurable with our current sensitivity. One possible explanation for that discrepancy might be the fact that the electrostatic model of Cho and coworkers[70] neglects polarizability and has been parameterized based on quantum chemical calculations focusing on hydrogen bonding in water clusters around a solvated azido group only, while L66AHA and I20AHA are basically not solvated (Fig. 3.5) and the electric field in Eq. (3.1) originates mostly from intra-peptide contact with polar side-groups.

In order to explain the distinct response of L78AHA, which actually reflects a change of intensity of the AHA vibration rather than its frequency, we speculate here that it originates from an electronic interaction between the AHA label and the sulfonate groups of the azobenzene-moiety. As a proof of principle, preliminary gas-phase quantum chemistry calculations, using structures derived from the classical MD simulation, give evidence that this might indeed be the case (Fig. 3.7). We must concede that we did not do these calculations on a sample of possible contact structures, that the agreement with experiment is rather modest, and that a polarizable force field or QM/MM simulations would be needed to verify that mechanism. As the interaction of the azido group with the sulfonate group is related to a weak chemical bond, the effect might be amplified in a QM/MM simulation, thus increasing the stability and occurrence of contact structures. The same seems to happen also upon hydrogen bonding to water, as evidenced by the observation of Cho and coworkers[70,71] that QM/MM simulations are necessary to correctly descibe the water structure around the azido group, which is a prerequisite for predicting a blue shift upon solvation.

Despite the preliminary character, the current results nevertheless show that the intensity of the AHA vibration carries information about its surrounding that is complementary to the frequency position. This effect has not been described in literature to the best of our knowledge, and needs to be taken into account when analyzing difference spectra from IR labels that can undergo electronic interactions with neighboring

amino acids, such as azido and cyano groups. In a broader sense, it reminds one that in difference spectroscopy, band disappearances do not necessarily need to be an effect of a vanishing state population, but can as well stem from effects altering the transition dipole moment of the reporting label. In conclusion, we have found a new and unexpected mode of spectral response of an AHA label incorporated into a protein. This response stems from an electronic instead of an electrostatic interaction of the AHA azido group with the protein/water surrounding, and opens up new applications for the usage of artificial probes in biospectroscopy.

3.5 Supplementary Information

AHA parameterization procedure

Force field parameters for the azobenzene photoswitch and the attached cysteine side chains as well as for AHA labels were obtained with the Antechamber package.[205] For the determination of atomic charges, the structures of the switch with attached cysteine side chains in *cis* and *trans* conformations as well as the AHA labels were optimized on B3LYP/6-31G* level using the GAUSSIAN g09 program suite.[143] Atomic charges of the different conformers were then computed as Mulliken charges from HF/6-31G* single point calculations. Point charges for MD calculations were then obtained from multiconformational restrained electrostatic potential (RESP)[158] calculations. The Cβ partial charges of covalently attached Cys residues were constrained to the value given in the Amber99sb*ILDN force field.

AHA parameters

[AHA]

[atoms]

N	N	-0.41570	1
H	H	0.27190	2
CA	CT	0.07730	3
HA	H1	0.07950	4
CB	CT	-0.11720	5
HB1	H1	0.05338	6
HB2	H1	0.05338	7
CG	CT	0.15260	8
HG1	H1	0.06012	9
HG2	H1	0.06012	10
N1	Nah	-0.44965	11
N2	Nbh	0.20880	12
N3	Nch	-0.06395	13
C	C	0.59730	14
O	O	-0.56790	15

[bonds]

-C	N
N	H
N	CA
CA	HA
CA	CB
CA	C
CB	HB1
CB	HB2
CB	CG
CG	HG1
CG	HG2
CG	N1
N1	N2
N2	N3
C	O

[impropers]

-C	CA	N	H
CA	+N	C	O

[atomtypes]

Nah	7	14.01	0.0000	A	3.25000e-01	7.11280e-01
Nbh	7	14.01	0.0000	A	3.25000e-01	7.11280e-01
Nch	7	14.01	0.0000	A	3 .25000e-01	7.11280e-01

[bondtypes]

Nah	CT	1	0.14770	262590.0 ; AHA side chain GAFF
Nah	Nbh	1	0.12160	717470.0 ; AHA side chain GAFF
Nbh	Nch	1	0.11240	992300.0 ; AHA side chain GAFF

[angletypes]
[dihedraltypes]

CT	CT	Nah	1	109.500	555.640 ; CB-CG1-N1 AHA GAFF
CT	Nah	Nbh	1	116.200	574.380 ;CG1-N1-N2 AHA GAFF
Nah	CT	H1	1	109.600	412.210 ; N1-CG1-HG12 AHA GAFF
Nah	Nbh	Nch	1	180.000	566.100 ; N1-N2-N3 AHA GAFF

CT	CT	CT	Nah	3	0.650	1.952	0.000	-2.603	0.000	0.000
CT	CT	Nah	Nbh	3	0.000	0.000	0.000	0.000	0.000	0.000
CT	Nah	Nbh	Nch	3	0.000	0.000	0.000	0.000	0.000	0.000
Nah	CT	CT	H1	3	0.650	1.952	0.000	-2.603	0.000	0.000
Nbh	Nah	CT	H1	3	0.000	0.000	0.000	0.000	0.000	0.000

Supplementary Figures

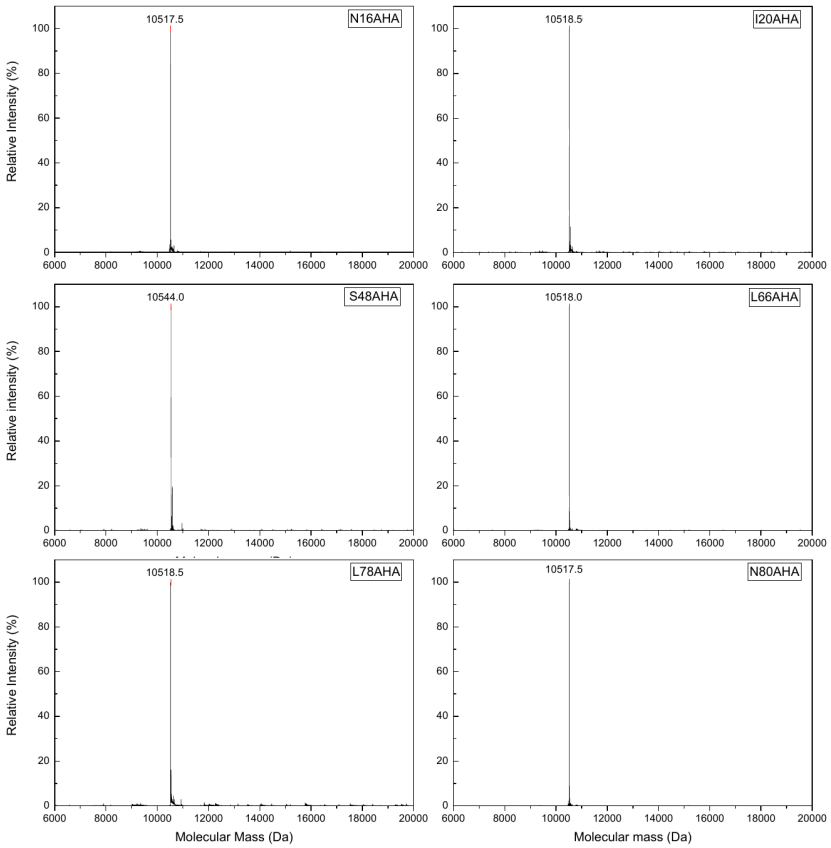


FIGURE S1: ESI-mass spectra of all mutants considered in this paper, evidencing the the excellent purity of the labelled and cross-linked protein.

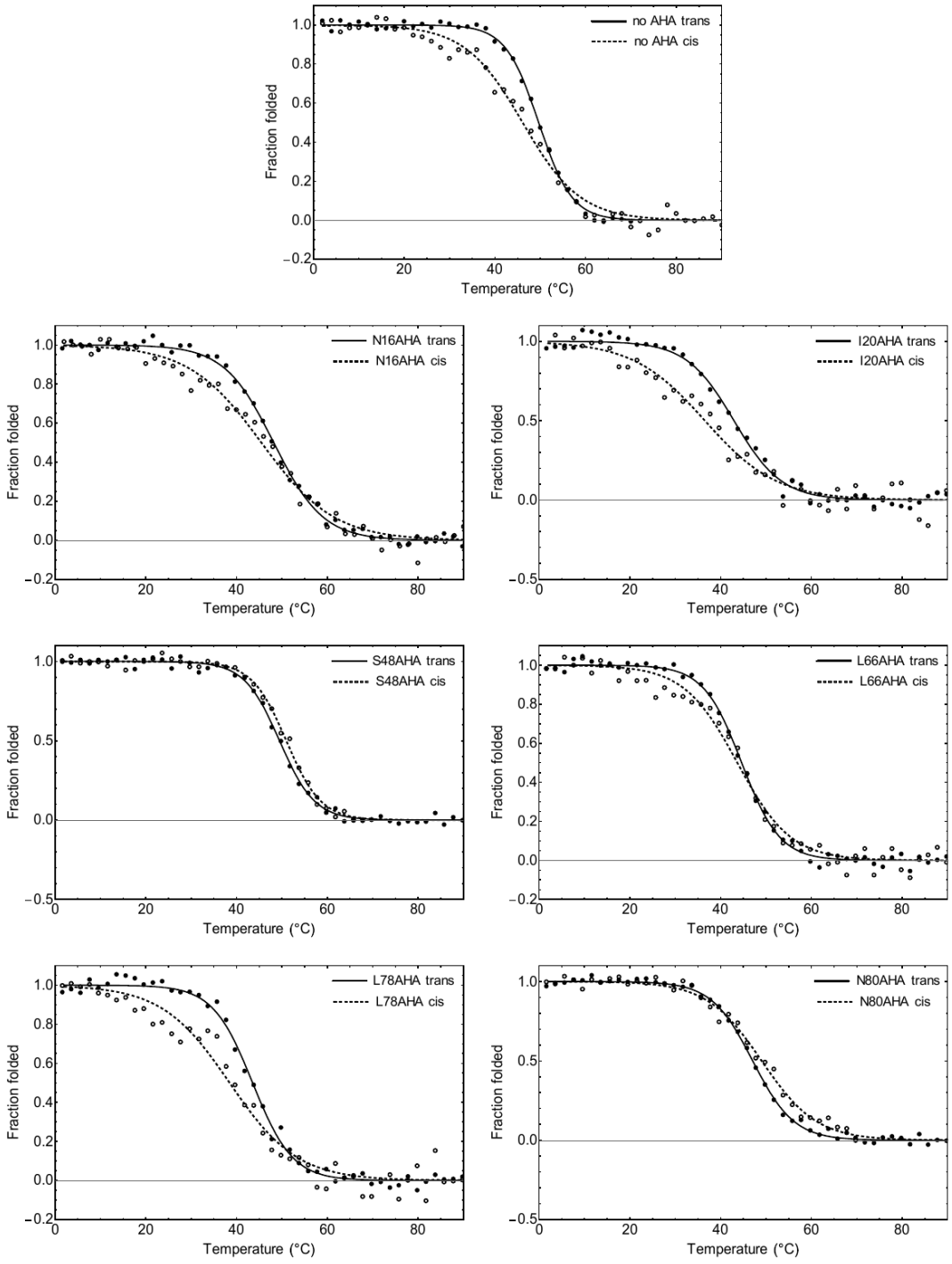


FIGURE S2: Temperature-induced unfolding measured by CD in the two states of the photo-switch of the PDZ2 domain without AHA label (top), and all the mutants considered in this paper. The *trans* data have been measured for the dark-adapted protein, the *cis* data after illumination for 2 min at 370 nm from a cw diode laser (CrystaLaser CL-2000). The data have been fit to a function $1/(1+\exp((T-T_m)/\Delta T))$ after subtraction of the background and normalisation. The fit results are summarized in Table 1.

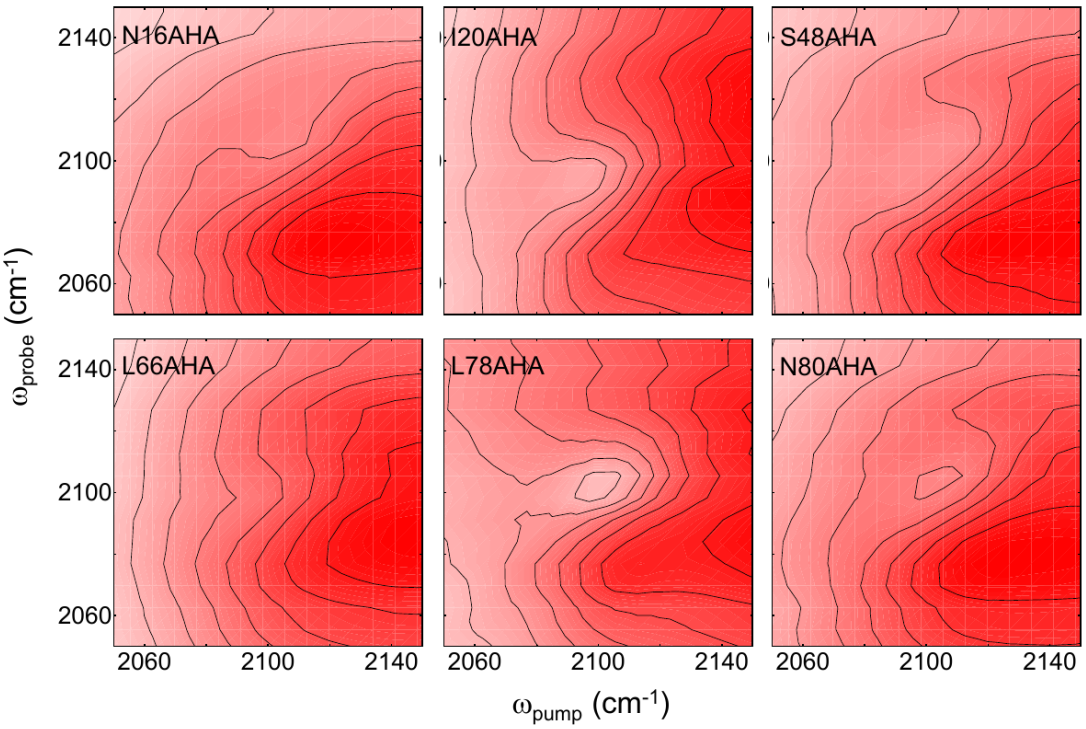


FIGURE S3: Purely absorptive 2D IR spectra of the dark-adapted *trans*-state of all mutants considered in this study. The AHA signal can be recognized, but is sitting on a huge water background (which has not been measured independently and subtracted out, since that would require much larger sample amounts). The water background is positive (red, excited state absorption), while the AHA diagonal band is seen as negative dip. The relative contribution of the AHA varies, since the concentration varied between 0.6 mM and 1.1 mM, given by the solubility of the particular mutant.

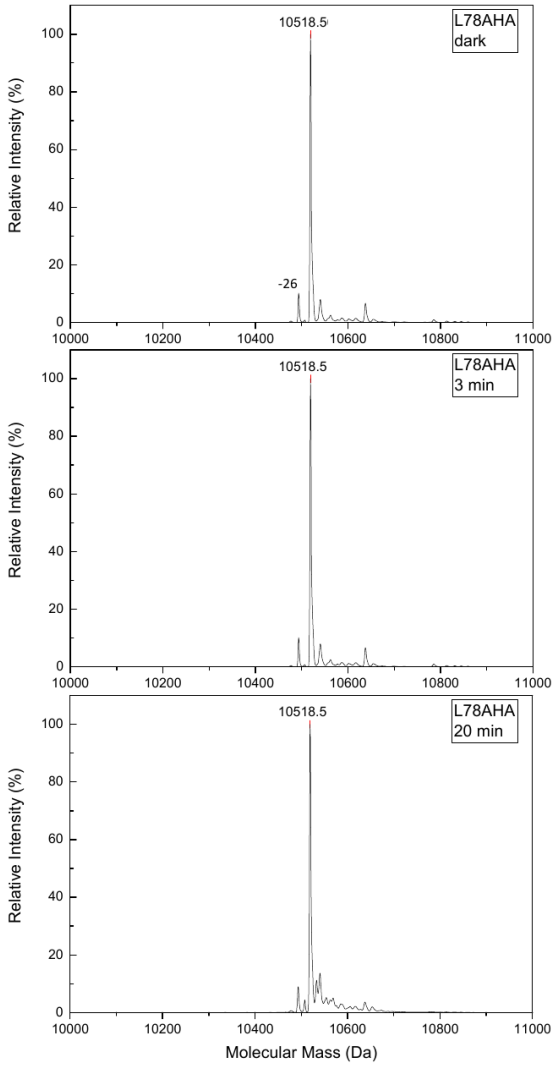


FIGURE S4: ESI-mass spectra of L78AHA (from a different batch than in Fig. S1) before illumination (top), and after 3 min (middle) or 20 min (bottom) of illumination at 370 nm from a cw diode laser (CrystaLaser CL-2000). The first case coincides with the conditions used to measure the 2D IR difference spectra of Fig. 3, while the second case applied significantly more light. Even in the second case, still basically no changes in the chemical composition of the sample are detectable, in particular no reduction of the azido-group that would lower the mass by 26 Da due to the loss of a N_2 molecule. The corresponding peak is labelled in the top panel, is very small, and does not increase upon illumination.

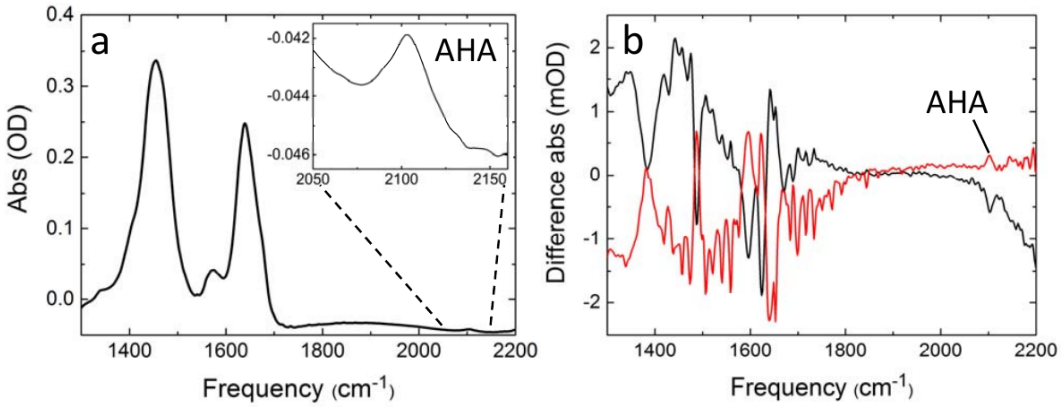


FIGURE S5: (a) Absolute FTIR spectrum of L78AHA with the buffer background subtracted; the inset focusing into the small AHA band. (b) Difference FTIR spectra of L78AHA. The black line shows the result when taking a background for the dark-adapted protein in *trans*, and then switching to *cis* by illumination at 370 nm. The dip at ≈2100 cm^{-1} indicates the loss of intensity of the AHA band. The red line shows the result when taking the *cis* configuration as background and then (partially) switching back to *trans* by illumination at 420 nm. The AHA band regains its intensity. Back-switching from *cis* to *trans* is not complete, since both the *cis* and the *trans* configuration absorb at 420 nm, thus establishing a photo-equilibrium (in contrast to *trans-cis* switching at 370 nm that can be performed to almost 100%). The results nevertheless show that back-switching is reversible.

Azidohomoalanine: A Minimally Invasive, Versatile, and Sensitive Infrared Label in Proteins To Study Ligand Binding
Here, I performed all MD simulations and analyzed the simulation data.

3.6 AHA Label as a minimally-invasive, versatile and sensitive Infrared probe.

Proteins are key mediators in virtually every biological process exhibiting their functions through specific interactions with other proteins, peptides, nucleic acids, small molecules, or ions.[3,214,215] Vital information in the cell is transmitted through specific non-covalent interactions in *e.g.* signal transduction cascades, regulation of metabolic processes, activation or inhibition of enzymatic reactions, assembly of macromolecular complexes, and programmed cell death.[96,216] Disruption of specific interactions is a major cause of numerous diseases. Therefore, understanding structural details of protein-ligand interactions is a prerequisite for clarifying cellular processes at a molecular level.[217] Furthermore, knowledge of the mechanisms of protein-ligand recognition facilitates rational drug design for treatments of different diseases.[218–220] Because of its importance, a plethora of methods for protein-ligand binding analysis have been developed..[221–223] Specificity is one of the crucial features of protein-ligand interactions and therefore methods that can directly discriminate between specific and non-specific binding are advantageous.[224,225]

It has recently been demonstrated that the presence of an infrared label at the site of recognition between a protein and its binding partner can give site specific insight into the underlying mechanisms of how signaling proteins function.[93,94,226–228] Such studies require a special vibrational label that absorbs in the spectral window between $\approx$1700 cm^{-1} to $\approx$2800 cm^{-1} in order to discriminate it from a huge protein background.[88–92,229] Among the possible molecular groups that have been proposed in that regard,[171–173,177–181,226–228] we favor the noncanonical amino acid azidohomoalanine (Aha). Aha has a relatively high extinction coefficient ($\approx$300-400 $M^{-1}cm^{-1}$),[93] and it has been shown to be an environment-sensitive infrared probe, capable of sensing the polarity of its environment via the frequency of its vibrational transition[70,71,93,94,176,230] Additionally, this noncanonical amino acid can be incorporated into a protein in virtually any position via methionine-auxotrophic expression strategy.[144,182,183] All these characteristics make Aha one of the most promising IR labels.

For the most part, Aha has only been used to detect quite significant structural changes in protein structure. For example, unfolding of an Aha mutant has been investigated, where the effect has been the largest when the label is incorporated in the core of the protein, with a frequency blue shift of 19 cm^{-1} upon thermal denaturation.[176] In that case, not only the structure of the protein changes dramatically upon unfolding, but

the environment of the Aha label also changes from hydrophobic in the core of a protein to fully solvent exposed. With regard to ligand binding, on the other hand, the Aha label has been incorporated into peptide fragments derived from the Ras-associated guanine nucleotide exchange factor 2 (Ra-GEF2),[93,94] whose C-terminal domain interacts with the PDZ2 domain of human phosphatase 1E.[189] PDZ-domain containing proteins mediate a wide variety of signaling processes in diverse organisms and thus represent excellent model systems for peptide binding studies.[127] Upon binding to a PDZ2 domain, the peptide ligand undergoes a structural change that is comparable to protein folding, that is, the peptide is a solvent-exposed random coil in its unbound state, but structurally well defined in its bound state. Furthermore, depending on the position of the Aha label, it points into the rather hydrophobic binding pocket, and consequently, a frequency shift of similar size (15 cm^{-1}) has been observed.[94]

With the present paper, we set out to explore how far we can push the method. That is, rather than having the Aha label in the peptide ligand,[93,94] we incorporate it in the protein sequence near the binding pocket at position K38. This site was chosen since previous MD simulations have suggested frequent salt-bridges of the peptide ligand with K38 during the process of binding.[95] Hence, we expect at most a marginal structural change of the protein upon ligand binding,[114] but the ligand will still modulate the degree of solvation of the Aha label. We apply 2D IR spectroscopy[231] in a dedicated 2D IR spectrometer[91] in order to be sensitive enough to measure very small frequency shifts of the Aha vibration. For the purpose of this work, the advantage of 2D IR spectroscopy is its quadratic signal dependence on the absorption cross section, which to a significant extent suppresses the strong water background. We furthermore present molecular dynamics (MD) simulations aimed to provide microscopic information on the structural changes around the binding pocket, and the mechanisms of solvation of the Aha label.

We considered the three molecular systems shown in Fig. 3.8 in this study. As a reference sample, we measured and MD simulated the K38Aha mutant of apo-PDZ2 (denoted K38Aha throughout this paper, see Fig. 3.8a). We then bound either the wild-type peptide (RWAKSEAKENEQVSAV) to its binding groove (denoted K38Aha+wtPep, see Fig. 3.8b), or a mutant of that peptide (RWAKSEAKE<u>C</u>EQVS<u>C</u>V) with an apolar azobenzene moiety attached to the peptide via the two cysteines (denoted K38Aha+azoPep, see Fig. 3.8c). The azobenzene variant has originally been developed to photoswitch the binding affinity of the peptide ligand, which will be the topic of another publication. Here, we make use of the hydrophobicity of the azobenzene moiety, which might affect the degree of solvation of the Aha label (note that in contrast to our previous studies,[144,184,185] the azobenzene moiety does not contain any -SO_3^- groups). All experiments have been performed in the dark to ensure that the azobenzene moiety is in its *trans* configuration.

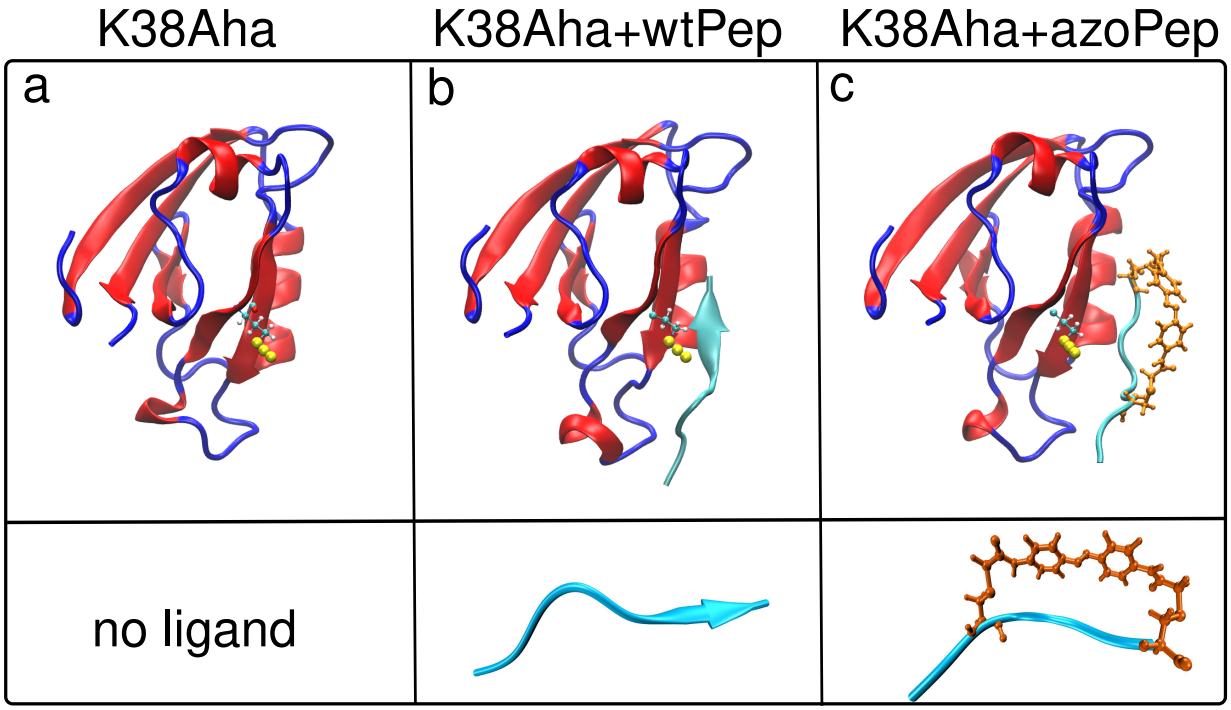


FIGURE 3.8: Molecular systems considered in this study. (a) The K38Aha mutant of apo-PDZ2 with the Aha label indicated in balls and sticks, (b) with the wild-type peptide bound to its binding groove (K38Aha+wtPep), and (c) with azobenzene-variant of the peptide bound to the binding groove (K38Aha+azoPep); azobenzene photoswitch displayed as orange sticks. The shown structures are snapshots taken from the MD simulation.

3.7 Methods

3.7.1 Protein and Peptide Preparation

Expression of the K38Aha mutant of PDZ2 domain was performed as described earlier.[144,183] The two peptide variants, wild type peptide (RWAKSEAKENEQVSAV) and its cysteine mutant ((RWAKSEAKECEQVSCV), were synthesized using solid phase peptide synthesis with standard 9-fluorenylmethoxycarbonyl (Fmoc) chemistry on a Liberty 1 peptide synthesizer (CEM Corporation, Matthews, NC, USA). The crude products were purified using reverse phase HPLC with a C18 column (Macherey Nagel, Dren, Germany), 0.1% TFA buffered acetonitrile gradient 0-50% in 10 column volumes.

The second peptide variant was linked to the azobenzene moiety (diiodoacetamide azobenzene) as reported previously[107] with the following modifications: 50 μmol of peptide was dissolved in 75 mL of 20 mM TRIS buffer, pH 8.5, which was previously extensively degassed. In order to promote the linking reaction by reducing disulphide bridges between cysteines, one equivalent of TCEP was added to peptide solution and allowed to incubate for one hour after which 100 μmol of azobenzene linker dissolved in 250 mL of THF was added. The solution was stirred and left in the dark at room temperature overnight. The linked peptide was concentrated at 30 °C to precipitate

the excess of apolar azobenzene linker. The suspension was filtered and the linked azo peptide was repurified using reverse phase HPLC with a C18 column.

All products, protein and both peptides, were dialyzed against 50 mM borate, 150 mM NaCl buffer, pH 8.5, lyophilized, and finally resuspended in H_2O for the circular dichroism (CD) and isothermal titration calorimetry (ITC) measurements and D_2O for the 2D IR measurements. Purity of all samples was confirmed by mass spectrometry analysis.

3.7.2 2D IR Spectroscopy

2D IR spectra in the boxcar geometry were measured using mid-IR pulses generated by a home-built optical parametric amplifier[198] providing ≈150 fs FWHM, ≈2 μJ pulses at ≈5 μm with a repetition rate of 5 kHz, as previously described.[94] A syringe pump sample delivery system was used to flow either sample or buffer through a cell comprised of two CaF_2 windows (2 mm thickness, 25 mm diameter) with a 25 μm teflon spacer. 2D IR spectra of the buffer and of the sample solutions were measured independently, exchanging approximately 250-300 μL of sample per measurement with the help of the syringe pumps. Buffer and sample spectra were phased individually, using the water background for a reference phase as described previously,[202] and then subtracted to reveal the isolated Aha signal.

3.7.3 Computational Methods

The starting structure of PDZ2 was taken from PDB entry 3LNX. To generate initial structures for the MD simulations of K38Aha+wtPep, we extracted the peptide structure and position from the X-ray structure 3LNY,[188] which contains an only 6 residues long peptide ligand, and attached 3 additional residues from the NMR structure of 1D5G[114] at the N-terminus. The peptide considered in the MD simulation (KENEQVSAV) thus is shorter than in experiment. K38 was mutated to Aha in all systems.

The MD equilibrium simulations with a production run length of 100 ns were performed using the GROMACS software package v2016.3[48] with a hybrid GPU-CPU acceleration scheme and the AMBER99SB-*ILDN forcefields as described in ref..[144] Azoswitch parameters were obtained using Antechamber[205] with BCC charges[232,233] and are described in the Supplementary Materials. We used a dodecahedral box with an image (i.e., face-to-face) distance of 7 nm. The box contained 7495 (K38Aha), 7997 (K38Aha+wtPep) and 10168 (K38Aha+azoPep) water molecules, respectively. Further simulation details can be found in Ref..[144]

To determine -N_3 contacts with the protein and the ligand, *g_mindist* from the Gromacs tools was employed. Contact distributions were then obtained by histogramming the MD data with 0.01 nm binning width. We define a contact to be formed if the minimal distance between azido group nitrogen atoms and protein/ligand residue is shorter

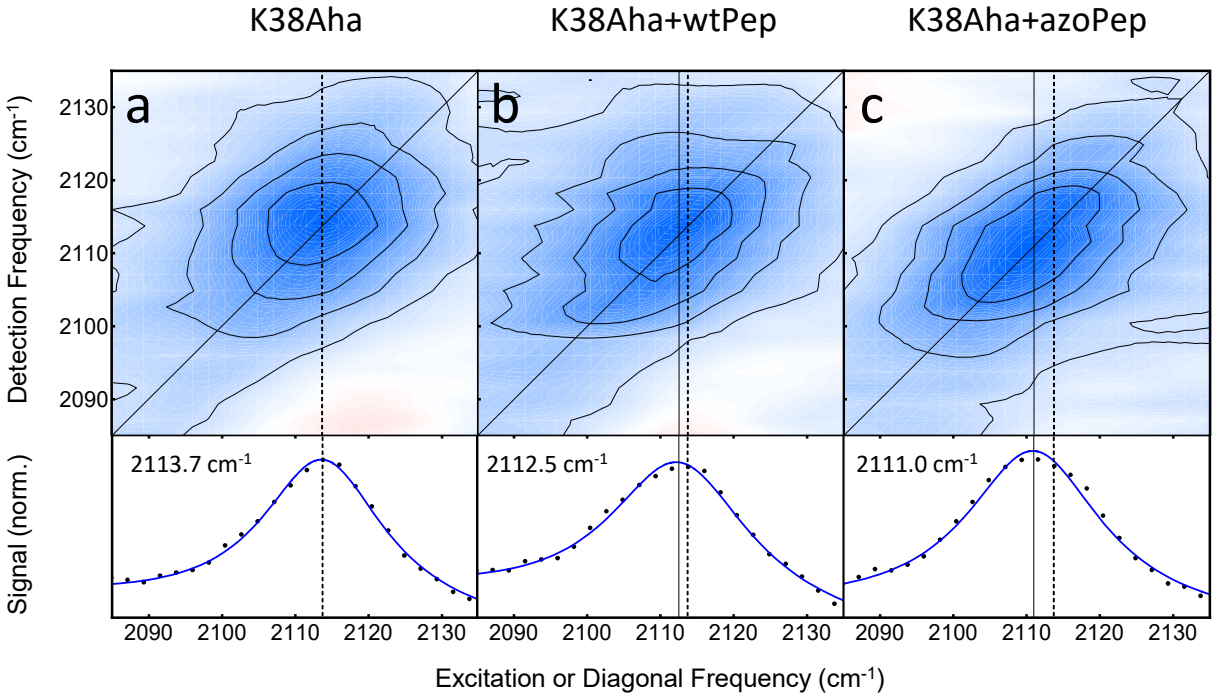


FIGURE 3.9: Ligand binding observed with 2D IR spectroscopy. (a) 2D IR response of K38Aha, (b) K38Aha+wtPep and (c) K38Aha+azoPep. The bottom panels plot the diagonal signal together with fits used to deduce the peak position (see text for details). In panel (a) the dotted line marks the reference peak position of the Aha label without any ligand. In panels (b+c), the peak positions of the Aha label with the two different ligands are marked in addition to it as solid lines.

than 0.45 nm.[52,234] In a similar way, we analyzed contacts between Aha and water as azido group/water oxygen atom distances with a cutoff of 0.45 nm.

We used the empirical model of Cho and co-workers[70] to estimate vibrational frequency shifts $\delta\omega$ caused by changes in the electrostatic environment of the Aha labels. By calculating the electric field $E_j(t)$ at the nitrogen atoms ($j = 1, 2, 3$) of the azido group for each MD snapshot at time t, we obtain the spectral shift (relative to the vacuum value)

$$\delta\omega(t) = \sum_j a_j E_j(t) \tag{3.2}$$

with coefficients a_j given in Ref..[70] Electric fields were computed via a reaction field approach using a cutoff radius r_c = 2.3 nm as described in Ref..[144] From the frequency trajectory $\delta\omega(t)$ with a time step of 15 ps, the distribution of the vibrational shifts was obtained via a histogram using 50 bins between -25 and 25 cm^{-1}.

3.8 Results

3.8.1 Experimental Results

To set the stage, we start with verifying that the Aha label only minimally disturbs the protein system. To that end, temperature dependent CD signals were recorded at 205 nm, revealing a melting temperatures of $44 \pm 1°C$ and $46 \pm 1°C$ for wild-type PDZ2 and K38Aha (Fig. S1, Supplementary Material), respectively. Hence, the mutation to Aha in fact slightly stabilizes the protein. Moreover, the dissociation constants of the wild-type peptide to wild-type protein and K38Aha are virtually the same (K_D=11 μM, see Fig. S2a,b, Supplementary Material). This is the evidence of the fact that the energy of binding has not been affected by substituting Aha in the vicinity of the binding groove of the protein. We also measured the binding affinity of the azo-peptide to K38Aha, revealing a somewhat smaller value (K_D=3 μM, see Fig. S2c, Supplementary Material) as compared to the wild-type peptide, indicating that the azo-moiety points away from the binding groove when bound to the protein, but capable of interacting with the protein in a stabilizing manner.

Fig. 3.9, top panels, show 2D IR spectra of the Aha vibrational label of all three sample systems (we concentrate here on the ground-state (0-1) peak of the 2D IR response on the diagonal, while the excited state (1-2) peak is outside the frequency window shown in Fig. 3.9). The concentrations used in these experiments were 0.9 mM for the protein, and 2.5 mM and 1.5 mM for wildtype and the azo-peptide, respectively. Given the binding affinities of both peptides in the low μM range (see Supplementary Material, Fig. S2), practically 100% of the protein, which carries the Aha label, has a ligand bound at these concentrations. Fig. 3.9, bottom panels show diagonal cuts through the 2D-IR spectra (averaging over the diagonal and the two first off-diagonals). To determine the peak frequency of the label, we fit these diagonal cuts to a line-shape function. We found that the quality of the fit is better when using a Lorentzian rather than a Gaussian line-shape, and we added a tilted linear baseline to the fit, which accounts for a residual water background. We considered the center of the fitted Lorentzian to be the peak position of the corresponding peak band. That procedure revealed a peak position of 2113.7 cm^{-1} for K38Aha (marked as dotted vertical lines in Fig. 3.9 a,b,c), and red shifted peaks at 2112.5 cm^{-1} for K38Aha+wtPep and 2111.0 cm^{-1} for K38Aha+azoPep (solid vertical lines in Fig. 3.9 b,c). In K38Aha, the linewidth is 20.5 cm^{-1} FWHM together with a rather round 2D IR lineshape (justifying the Lorentzian fit). Upon binding of a ligand, the linewidth increases a little bit (23.6 cm^{-1} for K38Aha+wtPep and 22.6 cm^{-1} for K38Aha+azoPep) and the 2D IR lineshape becomes more elongated along the diagonal, indicating a more inhomogeneously broadened absorption band.

By repeating these experiments various times (i.e., between 2-4 times for different samples), taking into account the uncertainty in the phasing,,[202] and by fitting the results to different models (e.g. fitting only 1D diagonal cuts as in Fig. 3.9 vs fitting the

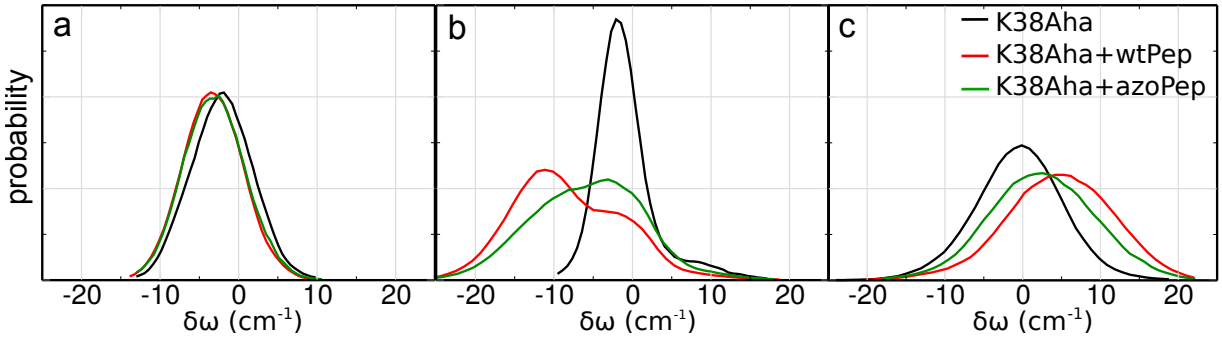


FIGURE 3.10: Vibrational frequency shifts (relative to the vacuum value) of all simulated systems by considering (a) all atoms, (b) only protein/ligand atoms and (c) only solvent atoms.

complete 2D lineshape or using different line-shape functions), we estimated that the statistical and systematic error in the determination of the peak-position of the Aha band is in the order of 0.5 cm^{-1}. One example of a completely different data set of K38Aha and K38Aha+wtPep is shown in Fig. S3, which has been measured with different concentration on a different day by different people. Applying the same fit function, the deduced peak frequencies are the same within 0.1 cm^{-1} as in Fig. 3.9, evidencing the reproducibility of the measurement. To conclude this part, for both K38Aha+wtPep and K38Aha+azoPep, we are confident that the effect is real with frequency red shifts of 1.2 ± 0.5 cm^{-1} and 2.7 ± 0.5 cm^{-1}, respectively.

3.8.2 Computational Results

Recent quantum-chemical calculations of Cho and coworkers[70,71] have shown that the spectroscopic signatures of the azido stretch mode of Aha mainly reflects the local electrostatic environment of the -N_3 reporter group. The structures shown in Fig. 3.8 suggest that the azido group is indeed found in different environments in each investigated case. That is, Aha is fully exposed to the solvent in K38Aha, and rarely interacts with protein atoms, while in the cases of K38Aha+wtPep and K38Aha+azoPep it is less exposed to the solvent, but interacts mostly with ligand residues. These differences are expected to be detectable in the vibrational spectra.

Indeed, using Equation (3.2) for calculating spectral shifts displayed in Fig. 3.10a, we find that the reporter group in K38Aha+wtPep and K38Aha+azoPep induces a red shift of 1-2 cm^{-1} with respect to K38Aha. To learn about the contribution of protein and solvent to the total vibrational frequency, we calculated the shifts due to protein/ligand atoms and solvent atoms independently. Considering only protein/ligand atoms, K38Aha+wtPep and K38Aha+azoPep show red shifts of 8-10 cm^{-1} and 1-3 cm^{-1} compared to K38Aha (Fig. 3.10b). Regarding solvent molecules only, K38Aha+wtPep and K38Aha+azoPep are blue shifted by 2-5 cm^{-1} (Fig. 3.10c).

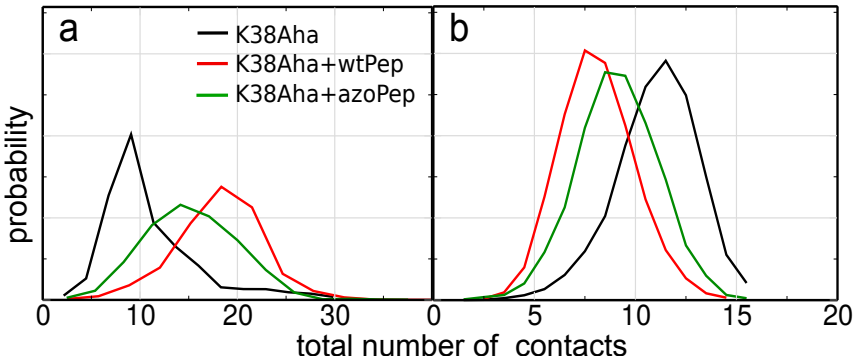


FIGURE 3.11: Probability distribution of average number of contacts of the azido group with (a) protein/ligand and (b) solvent molecules. The black curve shows the result for K38Aha, and the red and green curves that of K38Aha+wtPep and K38Aha+azoPep, respectively.

To relate the vibrational shifts to changes in contact patterns, Fig. 3.11 compares the probability distribution of the number of azido contacts with protein/ligand and water molecules. Concerning protein/ligand contacts, K38Aha+wtPep and K38Aha+azoPep simulations exhibit 12-18 protein/ligand contacts compared to K38Aha with 9 contacts. Regarding water contacts, the azido group of K38Aha+wtPep and K38Aha+azoPep exhibit 7-8 water contacts, while K38Aha on average forms 10-12 contacts with water molecules. Hence, water contacts are replaced by ligand contacts upon binding of the peptide ligand.

To further explore the cause of red shifts of K38Aha+wtPep and K38Aha+azoPep, we calculated minimum distances between the azido group and all residues of the respective systems. We found that significant contacts only appear with ligand residues (Fig. S4), as suggested in Ref..[95] From the minimum distance analysis, three major contacts of Aha38 with ligand residues Glu(-5), Gln(-4) and Val(-3) were identified as potential candidates to induce red shifts. However, the distance patterns of Gln(-4) and Val(-3) do not significantly change between both protein/ligand systems, while Glu(-5) is special, as it is the only one of the three residues with significant changes in distances. That is, the middle nitrogen atom of the azido group of K38Aha+wtPep forms a consistent polar contact with the negatively charged side chain carboxyl oxygen of Glu(-5) (Figure 3.12a), while this contact is less stable in K38Aha+azoPep. To test if the contact with Glu(-5) is the main origin of red shifts observed in experiments and simulations, we excluded this specific residue and compared the resulting spectrum with spectra including Glu(-5) (Fig. 3.12b,c). Exclusion of Glu(-5) induces a blue shift of $\sim$4-6 cm^{-1}, with a difference of $\sim$2-3 cm^{-1} between K38Aha+wtPep and K38Aha+azoPep. Very likely, it is these very specific, but fluctuating contacts that give rise to the larger inhomogeneity observed experimentally upon ligand binding in the form of 2D IR line shapes that are more elongated along the diagonal (Fig. 3.9 b,c).

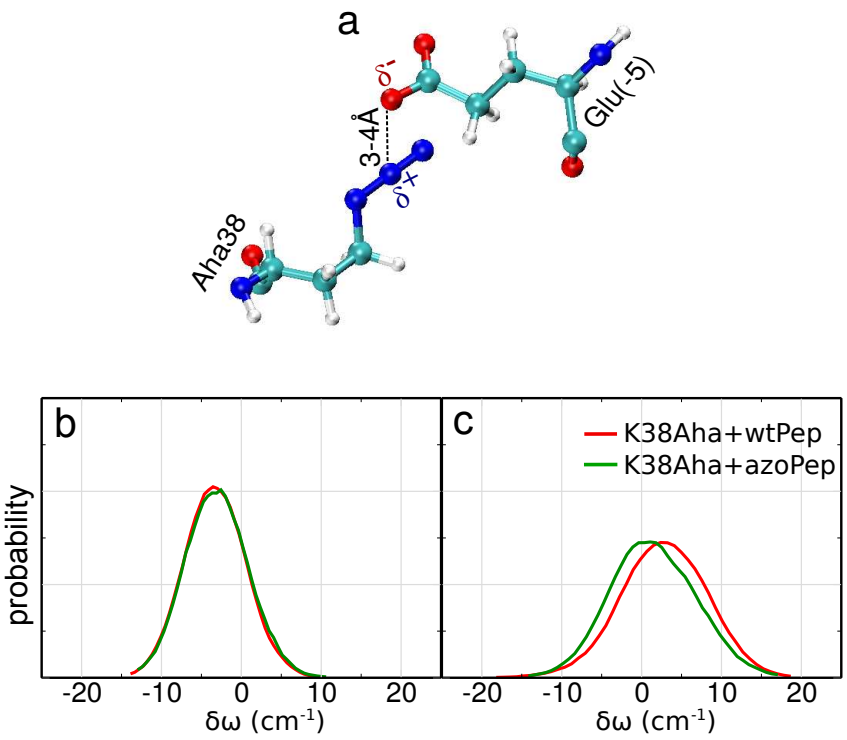


FIGURE 3.12: (a) Polar contact between the positively charged $N^{(2)}$ of the azido group and the negatively charged side chain carboxyl oxygen of Glu(-5). (b,c) Vibrational frequency shifts of K38Aha+wtPep and K38Aha+azoPep by including (b) and excluding (c) Glu(-5).

We also looked into the possible contribution of the ions in the solution to the frequency shift. Fig. S5 shows that the average distance between the azido group to the nearest ion is between 1.0 to 1.5 nm, and contacts (i.e., distances of less than 4.5 Å) almost never occur. We therefore conclude that the effect of the ions on vibrational frequency of the azido group is negligible due to the shielding by water.

3.9 Discussion and Conclusion

We have demonstrated that it is possible to reliably measure frequency shifts of the Aha label as small as ≈1 cm^{-1} with the help of 2D IR spectroscopy, sufficient to detect rather subtle structural changes of a protein system. Furthermore, MD simulation together with pre-parameterized spectroscopic maps[70] can qualitatively reproduce the size of the spectroscopic effect. However, while the experiments reveal a bigger effect for K38Aha+azoPep than for K38Aha+wtPep, consistent with the notion that the hydrophobic azobenzene moiety might shield the Aha label more efficiently from the solvent, that effect is not reproduced by the MD simulations. On the contrary, the MD simulations suggest that interactions with very specific side chains (in this case Glu(-5)) might have a dominating effect, the exact calculation of which would require more sophisticated QM/MM simulations.[70]

It is the common notion that the frequency shift of the azido group is a measure of the amount of solvation; the stronger an azido group is solvated, the more blue is its absorption band.[70,71,93,94,176,178,230] Our experimental observation supports that view, with the ligands shielding the Aha label to a certain extent from water (Fig. 3.9). The MD results of Fig. 3.11b indeed show that this is the case: the number of solvent contacts is reduced once a ligand is bound to the protein. Yet, Figs. 3.10, 3.11a and 3.12 suggest that that this common notion needs to be re-interpreted somewhat. That is, a smaller number of water contacts is inevitably compensated by a larger number of protein contacts (Fig. 3.11a vs Fig. 3.11b), which can be polar as well; in this case in particular due to Glu(-5) (Fig. 3.12). These protein contacts also reveal frequency shifts due to electrostatic interactions according to Eq. (3.2). In fact, since the protein is more structured than the solvent, these protein contacts are more stable on average and may overcompensate the contribution of the water in terms of the frequency shift (Fig. 3.10). In other words, Fig. 3.10 reveals that all polar interactions tend to produce a frequency red shift, in contrast to the common notion, and that reduced solvation by water is overcompensated by even stronger "solvation" from the protein. It should however be noted that this interpretation is at odds with the conclusions of Ref.,[178] which measured the frequency of isolated azido-groups also in very apolar solvents such as hexane, which might be assumed to be close to the situation in vacuum, but where the frequency is red shifted relative to that in water. However, in the present case, assuming the protein to be a hydrophobic bulk turns out not to be a viable approximation. Instead, especially due to the presence of Glu(-5), the protein resembles best a highly anisotropic polar "solvent" for the Aha azido group with stable charge distribution, resulting in the observed additional red shifts. It is interesting to compare the present results to those of previous work, where we used Aha labels at various positions to detect structural changes in a PDZ2 domain that has been made photo-switchable by covalently linking a similar azo-moiety across its binding groove.[144] Only L78Aha, which sat very close to the photoswitch in the binding groove, showed an effect; in that case a change in absorption intensity rather than frequency. Also for that example we concluded that the effect is related to very specific interactions of the Aha label with the azo-photoswitch. Previous work[93,94] has shown that Aha can be used as a reporter to determine biomolecular recognition when incorporated into the peptide ligand directly. In this study, Aha has instead been incorporated in the protein in the vicinity of the binding groove. In contrast to the ligand-incorporated reporter, we have shown by CD and ITC experiments that the presence of the Aha label neither perturbs the stability of the PDZ2 domain, nor its binding affinity for the peptide ligand, in agreement with recent NMR and X-ray experiments that came to the same conclusion for a similar molecular system.[155] The 2D IR spectra of Fig. 3.9 showed that Aha is indeed a sensitive probe of ligand binding, exhibiting a few wavenumbers red shift upon binding of two different peptides.

MD simulations can reproduce the effect essentially quantitatively, providing an atomistic picture of the interactions of the Aha label at the protein surface. It is important to stress that the center frequency of an isolated vibrational transitions, such as that of the Aha label, can be measured with an accuracy much better than its linewidth or than the spectral resolution of the 2D IR instrument. This is the equivalent of the very idea of sub-diffraction microscopy, where the position of individual switchable fluorescing chromophores can be measured with a precision much better than the resolution of the microscope, provided that their point spread functions don't spatially overlap.[235] In conclusion, the present study demonstrates that Aha can be employed as a specific IR reporter not only for big changes of its chemical environment (e.g., protein folding/unfolding events), but also for very subtle changes of the electrostatic environment at the protein surface.

3.10 Supplementary Information

CIRCULAR DICHROISM SPECTROSCOPY

Circular dichroism spectra were collected using a Jasco J-810 spectropolarimeter in 1 mm quartz cuvette (Hellma Analytics, Quartz SUPRASIL) using 350 μL of protein solution. 20 μM protein solutions of both wild type and K38Aha PDZ2 domain were prepared in 50 mM borate, 150 mM NaCl, pH 8.5 buffer. The same buffer was used as a background. Temperature-dependent CD-responses Spectra were recorded at 205 nm and fit in order to reveal a melting temperature of 46 $\pm$ 1°C (Fig. S1, red), somewhat larger than the corresponding value from the wild-typeprotein 44 $\pm$ 1°C (Fig. S1, blue).

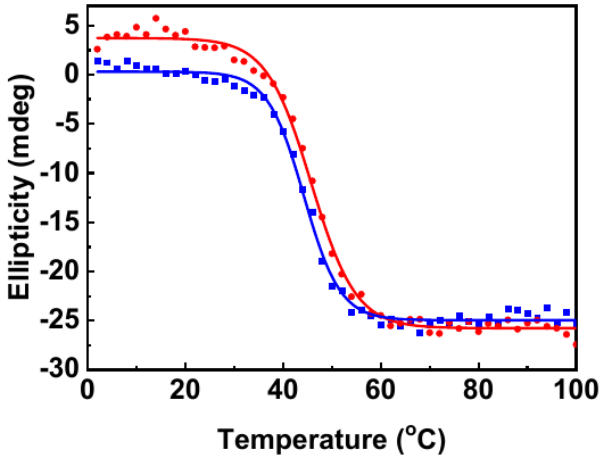


FIGURE S1: Normalized CD signal wildtype PDZ2 (blue) and the K38Aha mutant (red) recorded at 205 nm. The solid lines show two-state fits to the data, revealing melting temperatures of 44°C and 46°C, respectively.

ISOTHERMAL TITRATION CALORIMETRY

Isothermal titration calorimetry was performed using MicroCal^{TM} iTC200 system (Malvern Instruments Ltd, Malvern, UK) at 25°C (Fig. S2). Immediately before the experiment was performed, protein and peptide solutions were extensively redialyzed against 50 mM borate, 150 mM NaCl, pH 8.5 buffer. The sample cell of the calorimeter was loaded with 250 μL of 0.14 mM protein solution in 50 mM borate, 150 mM NaCl, pH 8.5 buffer. Circa 40 μL of 1.4 mM peptide solution in the same buffer was added in small increments (1.5 μL) with 120 seconds spacing between injections until the whole sigmoidal thermogram curve was obtained (26 injections). Stirring speed was set to 1000 rpm. Heats of dilution were determined by titrating the same peptide solution into the dialysis buffer and subtracted from the raw titration data before data analysis. The data

were processed using the Origin ITC software provided by the instrument manufacturer. Fitting was performed using one set of sites binding model and KD value was obtain directly from the fitted binding curve.

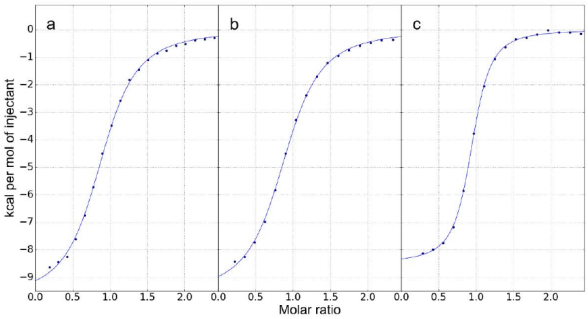


FIGURE S2: ITC measurements of the binding affinity of the wild type peptide to (a) the wildtype PDZ2 domain and (b) its K38Aha mutant. Panel (c) shows the result for the binding affinity of the azo-peptide to the K38Aha mutant of PDZ2. In either case, the concentration of protein has been 140 μM. The data have been fitted as explained in the text, revealing binding constants of KD =11 μM for (a) and (b), and KD =3 μM for (c).

INDEPENDENT DATA SET

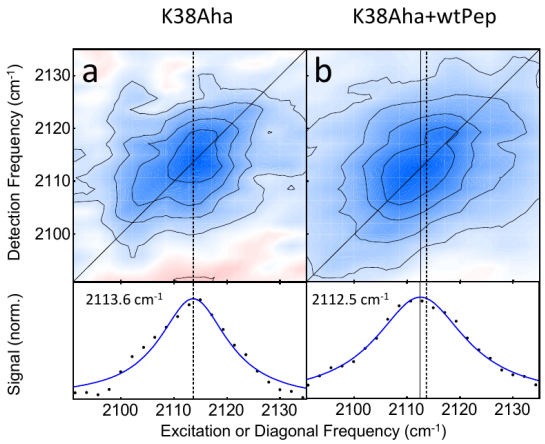


FIGURE S3: 2D IR response of K38Aha and (b) K38Aha+wtPep. In this case, the protein concentration has been 0.5 mM and the peptide concentration 2.5 mM. In panel (a) the dotted line marks the reference peak position of of K38Aha and inpanel (b) the peak positions of K38Aha+wtPep is marked in addition as solid lines. These data are, in principle, the same as in Fig. 2ab, but has been measured at a different day by different people. The deduced peak frequencies (2113.6 cm^{-1} and 2112.5 cm^{-1} , respectively) are the same within 0.1 cm^{-1} 1 as in Fig. 2ab, evidencing the reproducibility of these measurements.

MD CONTACT ANALYSIS

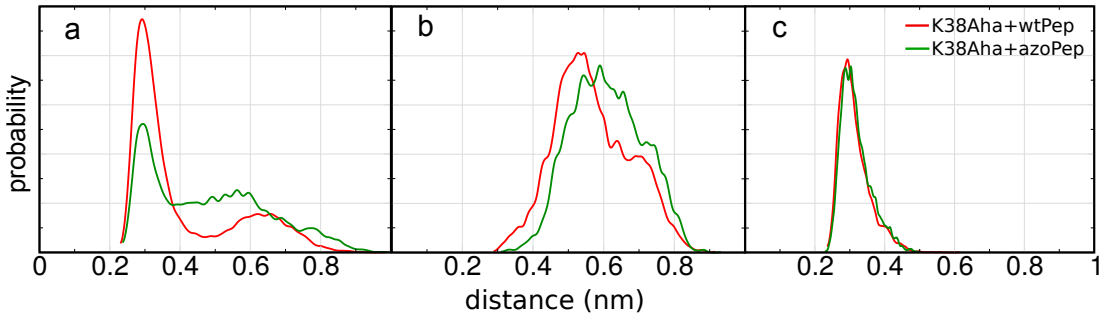


FIGURE S4: Probability distributions of minimum distances between ligand residues (a) Glu(-5), (b) Gln(-4) and (c) Val(-3) and the azido group of Aha38.

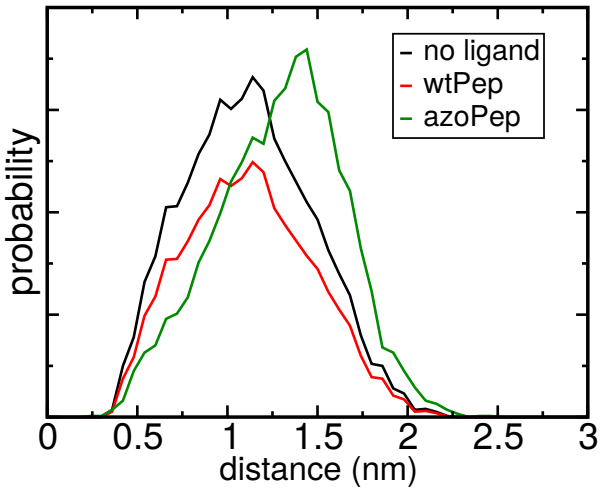


FIGURE S5: Probability distributions of minimum distances between the azido group of Aha38 and any ion in the simulation box.

AZO SWITCH PARAMETERS

Azo-switch parameters were obtained using Antechamber[205] and Acpype.[236] For parameterization, a model of the switch with attached cysteine side chains was built in Pymol[237] based on a preexisting structure with additional sulfonate side chains.[144] Atomic parameters were derived from GAFF parameters[142] with BCC charges,[232,233] the latter of which have been found to be consistent with the Amber protein force field used in the simulation.[238]

AZO Parameters

[AZO]

[atoms]

N	N	-0.41570	1
H	H	0.27190	2
CA	CT	0.02130	3
HA	H1	0.11240	4
CB	CT	-0.12310	5
HB1	H1	0.11120	6
HB2	H1	0.11120	7
SG	S	-0.05310	8
CD	CT	-0.32850	9
HD1	H1	0.15125	10
HD2	H1	0.15125	11
CE	C	0.30800	12
OE	O	-0.34000	13
NZ	N	-0.31550	14
HZ	H	0.24550	15
CH	CA	0.09400	16
CT1	CA	-0.18075	17
HT1	HA	0.15450	18
CI2	CA	-0.06850	19
HI2	HA	0.15100	20
CT2	CA	-0.18075	21
HT2	HA	0.15450	20
CI1	CA	-0.06850	26
HI1	HA	0.15100	27
CK1	CA	-0.07400	28
NL	NE	-0.07000	29
C	C	0.59730	30
O	O	-0.56790	31

[bonds]

N	H
N	CA
CA	HA
CA	CB
CA	C
CB	HB1
CB	HB2
CB	SG
SG	CD
CD	HD1
CD	HD2
CD	CE
CE	OE
CE	NZ
NZ	HZ
NZ	CH
CH	CT1
CH	CT2
CT1	HT1
CT1	CI1
CI1	HI1
CI1	CK1
CT2	HT2
CT2	CI2
CI2	HI2
CI2	CK1
CK1	NL
C	O
-C	N

[impropers]

-C	CA	N	H			
CA	+N	C	O			
N	CA	C	+N	105.4	0.75	1

[atomtypes]

NE	7	14.01	0.0000	A	3.25000e-01	7.11280e-01

[bondtypes]

NE	NE	1	0.12640	604239.0 ; AZO
NE	CA	1	0.14310	302956.8 ; AZO
CA	N	1	0.14220	311749.1 ; AZO

[angletypes]

NE	NE	CA	1	115.190	583.305 ; For the AZO NE NE CA
NE	CA	CA	1	119.880	567.228 ; For the AZO NE CA CA
CT	CA	CB	1	119.450	542.330 ; CB - CG - CD2 AZU GAFF
CA	CB	CA	1	114.190	570.360 ; CG - CD2 - CE2 AZU GAFF

[dihedraltypes]

CA	CA	N	H	9	180.0	1.88406	2 ; AZO side chain connection
NE	NE	CA	CA	9	180.0	0.00000	3 ; AZO side chain connection
CA	CA	N	C	9	180.0	1.88406	2 ; AZO side chain connection

Chapter 4

Photocontrolling Protein-Peptide Interactions

This chapter is based on following publication:
Photocontrolling Protein-Peptide Interactions: From Minimal Perturbation to Complete Unbinding
Brankica Jankovic[a,1], Adnan Gulzar[b,1], Claudio Zanobini[a], Olga Bozovic[a],Steffen Wolf[b] Gerhard Stock[b,1], Peter Hamm[a,1] *J. Am. Chem. Soc. 2019, 141, 10702-10710*

[a]*Department of Chemistry, University of Zurich, Zurich, Switzerland*
[b]*Biomolecular Dynamics, Institute of Physics, Albert Ludwigs University, Freiburg, Germany*
[1]*both authors contributed equally*
[*]*corresponding authors: peter.hamm@chem.uzh.ch and stock@physik.uni-freiburg.de*

An azobenzene-derived photoswitch has been covalently cross-linked to two sites of the S-peptide in the RNase S complex in a manner that the α-helical content of the S-peptide reduces upon *cis*-to-*trans* isomerization of the photoswitch. Three complementary experimental techniques have been employed, isothermal titration calorimetry (ITC), circular dichroism (CD) spectroscopy and intrinsic tyrosine fluorescence quenching, to determine the binding affinity of the S-peptide to the S-protein in the two states of the photoswitch. Five mutants with the photoswitch attached to different sites of the S-peptide have been explored, with the goal to maximize the change in binding affinity upon photoswitching, and to identify the mechanisms that determine the binding affinity. With regard to the first goal, one mutant has been identified, which binds with reasonable affinity in the one state of the photoswitch, while specific binding is completely switched off in the other state. With regard to the second goal, accompanying molecular dynamics (MD) simulations combined with a quantitative structure activity relationship (QSAR) revealed that the α-helicity of the S-peptide in the binding pocket correlates surprisingly well with measured dissociation constants. Moreover, the simulations show that both photo-switched configurations of all S-peptides exhibit quite well-defined structures even in apparently disordered states, which suggest an "induced fit" mechanism of binding.

Photocontrolling Protein-Peptide Interactions: From Minimal Perturbation to Complete Unbinding
Here, I performed all MD simulations and analyzed the simulation data.

4.1 Design of photoswitchable peptide

Understanding conformational and dynamic aspects of protein-protein interactions is of paramount importance for new strategies in controlling them. In order to employ time-resolved techniques for the investigation of conformational changes of proteins and/or peptides, e.g., by transient X-ray scattering[239] or transient IR spectroscopy,[240] fast triggering of a given process is required. This is where the importance for external control of specific interactions emerges - it facilitates detailed understanding of real-time conformational changes induced by protein-peptide recognition. Photocontrol has been recognized as a valuable tool for the external manipulation of numerous *in vitro* as well as *in vivo* processes. This way of triggering conformational changes offers numerous advantages, in particular high spatial and temporal resolution and selectivity.[98] Photoswitching, as one of the most promising approaches of photocontrol, has the additional advantage of being reversible.[97] For example, a reversible photocontrol of the α-helical content of small peptides with sufficient α-helical propensity can be achieved by crosslinking to sites of the peptide with an azobenzene-based molecule.[44,107–109] Upon illumination, the azobenzene chromophore undergoes a reversible *cis* to *trans* isomerization around its central diazene (N=N) double bond. Isomerization changes the geometry of the azobenzene moiety with different end-to-end distances of the two configurations, leading to either a perturbation or a stabilization of the helix, depending on the distance between sites to which the azobenzene moiety is bound.[99] Besides triggering folding/unfolding of small α-helices, azobenzene derivatives have also been incorporated in small cyclic peptides[241,242] and β-hairpins.[243–246] More complex structural processes induced by that concept include the folding/unfolding transition of relatively large proteins,[247,248] the modulation of enzymatic activity,[249–251] cell-cell adhesion,[252] or ion-channel activity.[98] *In vivo* applications became possible with the design of biocompatible azobenzene derivatives, e.g. in zebrafish embrios[253,254] or to control mitosis.[255]

In this study, we modulate the binding affinity within the non-covalent ribonuclease S complex (RNase S) with the ultimate goal of achieving photoinduced peptide unbinding. RNase S results from limited, site-specific hydrolysis of ribonuclease A, an enzyme from bovine pancreas.[100] Under controlled conditions, subtilisin can cleave a single peptide bond in RNase A and yields the RNase S complex composed of the S-peptide (residues 1-20) and the S-protein (residues 21-124).[100] Full enzymatic

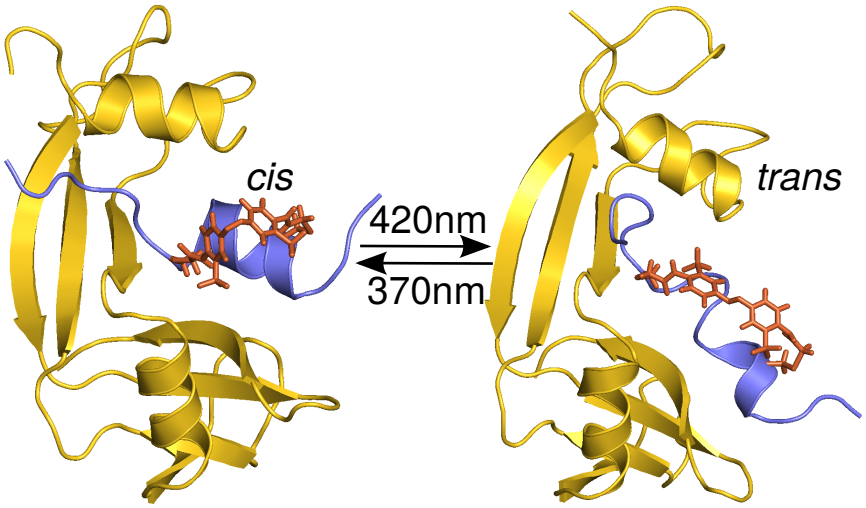


FIGURE 4.1: Molecular construct studied in this work: S-protein (yellow) with S-pep(6,13) (blue) bound to it in the two states of the photoswitch (red).

activity is restored when the two components form a native-like complex. The three-dimensional structure of the complex is essentially the same as that of intact RNase A, except from small differences near the cleaved peptide bond.[101] Consequently, the S-peptide adopts an α-helical structure when it is bound to the S-protein, while it is essentially disordered when isolated in solution due to its short sequence.[102] This property makes RNase S an excellent model to study the question whether the recognition mechanism between the S-protein and S-peptide can be characterized as "induced-fit" or as "conformational selection".[103–106] Furthermore, it opens the possibility to phototrigger binding/unbinding of the S-peptide, adopting the concept that has previously been used to control the α-helical content of isolated peptides via cross-linking two sites of the helix with an azobenzene moiety.[44,107–109] Fig. 4.1 shows the construct we designed for this work: a S-peptide (blue) is bound to a S-protein (yellow), the former of which being photoswitchable and designed in a way that the azobenzene-moiety (red) in its *trans*-configuration destabilizes the α-helical content of the S-peptide.

There have been attempts to photocontrol RNase S in the past, mainly with the focus on influencing the enzymatic activity. Pioneering studies have been performed by Lui et al.[256] as well as Hamachi et al.[257] They created different mutants with the non-natural amino acid phenylazophenylalanine placed at different positions of the S-peptide. In contrast to our construct, the photoswitch does not cross-link two sites. Nonetheless, by photoswitching the configuration of the side chain, particularly at position 13 of the S-peptide,[257] the enzymatic activity in one of the photoswitch states was reported to completely cease. In the subsequent study of Woolley's group,[258] the same model system has been studied more extensively. They could not reproduce the results from Ref.,[257] rather they found only a two-fold difference in the activity between *cis* and *trans*

states. In this study, binding affinities were also estimated, albeit indirectly via the enzymatic activity results. The largest modulation they found for the binding affinity in the *cis* vs the *trans* state was a factor 5. A different protein system, that however uses a very similar design idea as our present one, has been reported by Kneissl et al.,[259] who designed photocontrollable α-helical model peptides that bind to the Bcl-xl protein and are relevant for apoptosis. Up to 20-fold difference in binding affinity have been reported upon *trans* to *cis* isomerization of the photoswitch, but specific binding was preserved in both states with binding affinities in the nanomolar range.

4.2 Design Criteria

For the design of the photoswitchable S-peptides, the following basic criteria were considered:

- A water-soluble derivative of azobenzene has been used as photoswitch, BSBCA (3,3′-bis(sulfonato)-4,4′-bis(chloroacetamido) azobenzene).[115]
- A pair of amino acid residues needs to be replaced by cysteines for the site-selective linking of BSBCA. When choosing these sites, their effect for binding to the S-protein and/or the stability of the α-helix needs to be considered.
- Steric hindrance between crosslinked photoswitch and S-protein binding cleft should be avoided so that specific binding is preserved. To that end, the PyMOL software package[237] was used for visual examination and discrimination between solvent-exposed and non-solvent exposed amino acid residues of the S-peptide.
- The goal has been to stabilize the α-helical conformation in the more compact *cis* state of BSBCA, and to destabilize it in the *trans* state. This criterion determines the distance between sites that are being crosslinked.[99] To that end, we assumed an ideal length of BSBCA in its *cis* state of 11-16 Å,[248] and determined distances from the X-ray structure of RNase A (PDB code: 2E3W).[260]

Within that framework, five different photoswitchable S-peptide variants have been designed. Ala6 has been chosen as the first anchoring point in all but one S-peptide variants, since it is the first amino acid residue (starting from N-terminus) that is part of the α-helical region and at the same time is solvent-exposed in the complex. Furthermore, it has been shown to have no effect on the binding affinity or the structure of the RNase S complex.[103,106,261]

To obtain a prototype S-peptide that preserves binding, we started with Ser15 as second anchoring point, which also is known to be a residue that does not affect the binding,[103] and furthermore, it fulfills all aforementioned steric criteria; in particular, the end-to-end distance of the photoswitch in the *cis*-state fits the distance in the X-ray

structure the best. We will denote that variant as S-pep(6,15). However, we learned quickly that this particular variant does not modulate the binding affinity significantly upon photoswitching, and we moved towards the next S-peptide variant, investigating the effect of a mutation that destabilizes the S-protein/S-peptide complex. We selected His12 for this purpose, whose mutation to Phe12 decreases the binding affinity by about a factor 25;[104] we will call that variant S-pep(6,15)H12F.

Furthermore, assuming that the α-helical content determines the binding, we considered S-pep(6,13) and S-pep(6,10) with shorter spacings 7 and 4, respectively, between the anchoring groups. According to Ref.,[99] these spacings maximize the difference in stability of an α-helix with the photoswitch in the *cis* vs the *trans* states. Finally, we explored one variant in which we varied the starting point by one amino acid, S-pep(7,11), again with spacing 4. The amino acids that are affected in this case, Lys7 and Gln11, were both shown to play a role in the stabilization of the helix and in complex formation.[262,263]

4.3 Materials and Methods

4.3.1 Preparation and Purification of S-Protein and S-Peptide Variants

The S-protein was prepared from the commercial bovine ribonuclease A (Sigma Aldrich) by limited proteolysis with subtilisin, in essence as described in Refs..[100,179] The proteolysis mixture was then applied to size-exclusion chromatography on Sephadex G-75 in order to isolate the S-protein from the S-peptide as well as from subtilisin and residual unhydrolized RNase A. Wild-type S-peptide (sequence KETAAAKFERQHMDSSTSAA) and four modified S-peptide mutants with pairs of cysteine residues on different positions (i.e., S-pep(6,15), S-pep(6,13), S-pep(6,10) and S-pep(7,11), see Sec. 4.2) were synthesized by standard fluorenylmethoxycarbonyl(Fmoc)-based solid-phase peptide-synthesis on a Liberty 1 peptide synthesizer (CEM Corporation, Matthews, NC, USA). For S-pep(6,15)H12F, an additional His-to-Phe mutation was introduced at position 12. All amino acids were purchased from Novabiochem (La Jolla, CA, USA). The peptides were subsequently purified using reverse-phase C18 HPLC chromatography with an acetonitrile gradient 0-100% in 10 column volumes. The cysteine residues were subsequently cross-linked with the photoisomerizable linker BSBCA,[115] and purified again with the same conditions as in the previous step. The purity of S-protein and all S-peptide variants was confirmed by mass-spectrometry. Solutions of S-protein and S-peptide variants were prepared in 50 mM sodium phosphate buffer at pH 7.0. Concentrations have been determined by amino acid analysis.

The *trans*-configuration is the thermodynamically more stable one and accumulates to essentially 100% upon thermal relaxation, which happens on a timescale of 1-10 h (see Supporting Information, Fig. S1A). Typically, we left the sample in dark overnight

before an experiment of the *trans*-configuration was performed. To prepare the *cis*-configuration, we illuminated the sample with a 370 nm cw diode laser (90 mW, Crysta-Laser). Its power is sufficient to switch a sample in minutes, and we verified that >85% *cis* are prepared in that way (see Fig. S1B and its discussion).

4.3.2 Isothermal Titration Calorimetry

The isothermal titration calorimetry (ITC) measurements were performed on a MicroCal iTC200 from Malvern (Malvern, UK) at 25°C or 21°C for wild-type and photoswitchable S-peptides, respectively. The sample cell contained 250 μL of S-protein solution at concentrations that varied according to the expected affinity. 10x higher S-peptide concentration was titrated into the sample cell. The first injection was 0.4 μL and subsequent injections every 120 s were 2.0 μL. We verified that ITC reveals reliable results for the binding affinity also in the *cis*-state, despite the fact that the overall measurement time is comparable to that of the thermal *cis*-to-*trans* back-isomerization for the free peptide, and despite the fact that this reaction releases heat[264] that is comparable to the heat of binding (see Fig. S1A and its discussion). K_d's were determined from the ITC data using the instrument's software and assuming one binding site.

4.3.3 CD Measurements

Circular Dichroism (CD) measurements were performed on a Jasco (Easton, MD) model J-810 spectropolarimeter in a 0.1 cm quartz cuvette. We made sure the measurement was fast enough to prevent significant amount of thermal back-isomerization, and that the intensity of the light used by the CD spectrometer was low enough so the amount of molecules that photoisomerized due to that light is negligible (see Fig. S2 in SI).

4.3.4 Intrinsic Tyrosine Fluorescence Quenching

Fluorescence measurements were performed using two different fluorimeters. A commercial instrument (PerkinElmer) was used for the wild-type S-peptide, setting excitation and emission wavelengths to 285 nm and 306 nm, respectively. A larger sensitivity was needed for the photoswitchable variants due to the large absorption of the photoswitch itself, which furthermore changes between *cis* and *trans* states, and hence required extremely low concentration to avoid fluorescence re-absorption (optical densities were kept below OD=0.05). At the same time, the excitation intensity needed to be very low to avoid inducing isomerization of the photoswitch during the course of the measurement (see SI for details). The larger sensitivity was achieved with a home-built fluorimeter, that employed 266 nm for the excitation obtained from the 3rd harmonic of a Ti:Sapphire amplified laser system (Spectra Physics, Spitfire). The fluorescence was detected in a 90 degrees geometry using a single photon counter (PMA 175-N-M, PicoQuant) together with a special filter (XRR0340 Asahi Spectra USA) to remove both the

266 nm and the 370 nm radiations needed to promote the fluorescence of the tyrosine's and the *trans* to *cis* isomerization of the photoswitch, respectively.

4.3.5 MD Simulations

To generate initial structures of RNase S for the MD simulations, we started with the RNase complex structure (PDB ID 1Z3P),[265] mutated Nva13 to Met13, and then attached five additional residues from an X-ray structure of RNase A (PDB ID 2E3W)[260] at the C-terminus of the S-peptide. Following the procedure from Ref.,[159] we next attached the azobenzene photoswitch on cysteine side chains by mutating the respective residues i, j of S-peptides according to experimental labeling positions (see Sec. 4.2). All MD simulations were performed using the GROMACS software package v4.6.7 (Ref.[48]) with hybrid GPU-CPU acceleration scheme, the Amber03ws forcefield[266] and TIP4P-2005 water[267] solvent molecules. After energy minimization, all systems were equilibrated for 10 ns, and simulated at 300 K for at least 400 ns (see Supporting Information for details). Moreover, a MD simulation of wild type RNase S was performed, in order to show the similarity of the conformational distributions of crosslinked vs native peptides (Fig. S3).

MD trajectories were preprocessed by performing a principal component analysis on backbone dihedral angles (dPCA+)[268] of residues 3 to 13 of S-peptides. Including only the first few principal components with 80% of the total sum of eigenvalues (depending on the individual system, about five to six), we define a reduced coordinate space of significantly lower dimensionality.[53] After dimensionality reduction, robust density-based clustering[54] was performed to identify the metastable conformational states of the individual systems (see Supporting Information for details). On the basis of most populated states, representative structures and respective flexibilities were obtained by employing *gmx rmsf* and displayed via *PyMOL*.[237]

4.4 Results

We used three independent methods to determine the binding affinity of the various S-peptides to the S-protein, isothermal titration calorimetry (ITC), circular dichroism (CD) spectroscopy and intrinsic tyrosine fluorescence quenching. Each one of these methods focuses on different aspects of binding, and has its strengths and weaknesses. For some of the S-peptides, more than one method could be applied with reasonable confidence, allowing us to cross-validate the results.

Starting with the "gold standard", ITC,[269] it was possible to determine binding affinities for relatively strong binders, see Fig. 4.2 and Tab. 4.1. Due to the limited solubility of both the S-protein and the S-peptide, we could not determine the binding affinity of

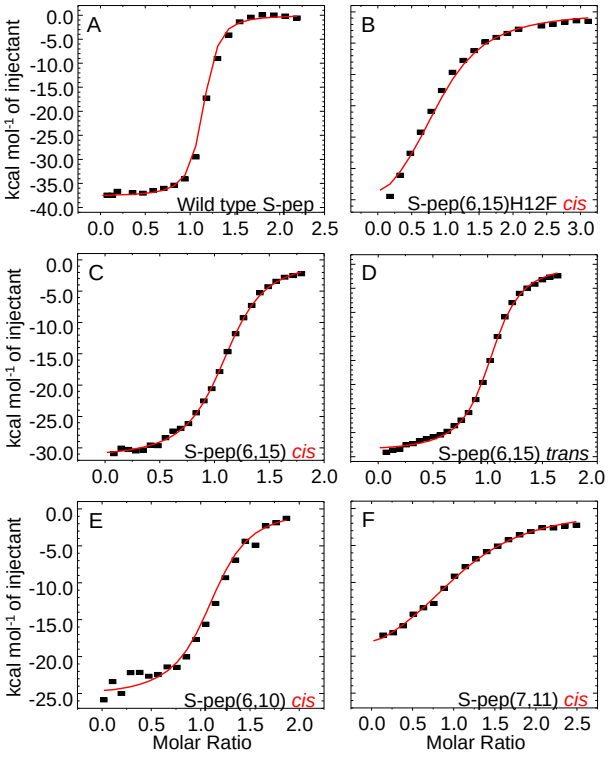


FIGURE 4.2: ITC thermograms of the S-protein and different S-peptide variants. (A) Wild-type S-peptide, (B) S-pep(6,15)H12F-*cis*, (C) S-pep(6,15)-*cis*, (D) S-pep(6,15)-*trans*, (E) S-pep(6,10)-*cis* and (F) S-pep(7,11)-*cis*.

S-pep(6,13) in either state of the photoswitch, and, more severely, also not for the *trans*-configuration of most of the other S-peptides, with the one exception of S-pep(6,15), for which the effect of switching is negligible.

We therefore turn to CD spectroscopy in the next step, as our starting assumption has been that the photoswitch in the *trans*-configuration perturbs the helical conformation of the S-peptide, and CD spectroscopy is sensitive to exactly that structural aspect. To set the stage, we show in Fig. 4.3 CD spectra of the S-pep(7,11) alone in its two states (red and blue), of the S-protein alone (black), as well as of mixtures of both at a concentration where the amount of binding has been ≈25% in the *trans*-state, and ≈75% in the *cis*-state (*vide infra*). For the S-peptide alone, the CD spectra indicate a predominantly random-coil conformation, regardless of the state of the photoswitch (see Fig. 4.3, red and blue). This finding is in agreement with the well-known fact that the S-peptide is disordered when it is free in solution.[104] The structural constrain of the photoswitch does not lead to any significant stabilization of the helical structure in the isolated form. Once bound to S-protein, however, the S-peptide adopts a helical structure, as deduced from a CD signal (Fig. 4.3, green and purple) that is significantly larger than the sum of the signals from the S-peptide alone (red and blue) and S-protein alone (black). Furthermore, the results show that there are significant differences in helical content between the *cis* and *trans* states, indicating a larger amount of binding due to a higher affinity in the *cis*-state.

TABLE 4.1: Dissociation constants K_d (in μM) of the various S-peptide variants and the S-protein, as determined by ITC, CD and intrinsic tyrosine fluorescence quenching.

	ITC		CD		fluorescence	
	cis	*trans*	*cis*	*trans*	*cis*	*trans*
S-pep(6,15)	1.5±0.2	1.2±0.8	1.7±1.0	2.9±0.8	3.6±0.9	3.2±1.3
S-pep(6,15)H12F	1.9±0.4	-	1.5±0.5	21±1.8	-	-
S-pep(6,13)	-	-	70±20	*1	130±30	*2
S-pep(6,10)	0.6±0.2	-	2.3±1.2	47±10	-	-
S-pep(7,11)	5.9±0.2	-	6.2±2.9	125±15	-	-
wild type	0.11±0.01		0.14±0.04		0.14±0.02	

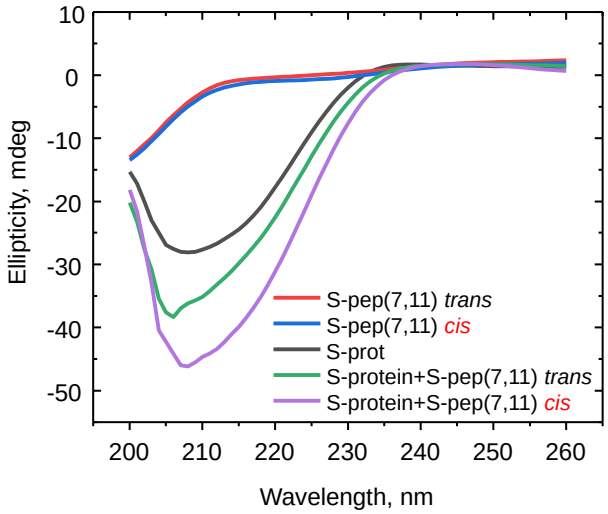


FIGURE 4.3: CD spectra of the S-peptide alone in *trans* (red) and *cis* (blue), the S-protein alone (black), and the S-protein with S-peptide in the *trans* (green) and *cis* state (purple), exemplified here for S-pep(7,11).

Besides folding of the S-peptide upon binding, the contribution of the S-protein *per se* to the overall CD signal might change as well. That can be seen from the fact that the difference between e.g. the purple line in Fig. 4.3 (S-protein plus ≈75% of bound S-pep(7,11)-*cis*) and the black line (S-protein alone) is much bigger than what the size of the fragments would suggest (20 amino acids for the S-peptide vs 104 amino acids of the S-protein). Indeed, it has been suggested that the S-protein is much more flexible in the absence of the S-peptide,[270] which is also the reason for difficulties in obtaining a crystal structure. Along these lines, a recent MD simulation has shown that helix II of the S-protein unfolds when the S-peptide unbinds.[106]

To determine the binding affinity, the CD signal at 225 nm[271] was measured as a function of S-peptide concentration, keeping the concentration of the S-protein constant at around 50 μM, see Fig. 4.4. In these plots, "0%-fraction-bound" corresponds to the CD signal calculated as a trivial sum of the contributions from S-peptide alone and S-protein alone, and "100%-fraction-bound" to the plateau value reached at high enough S-peptide concentrations. To determine the binding affinity K_d, the data were fit to a

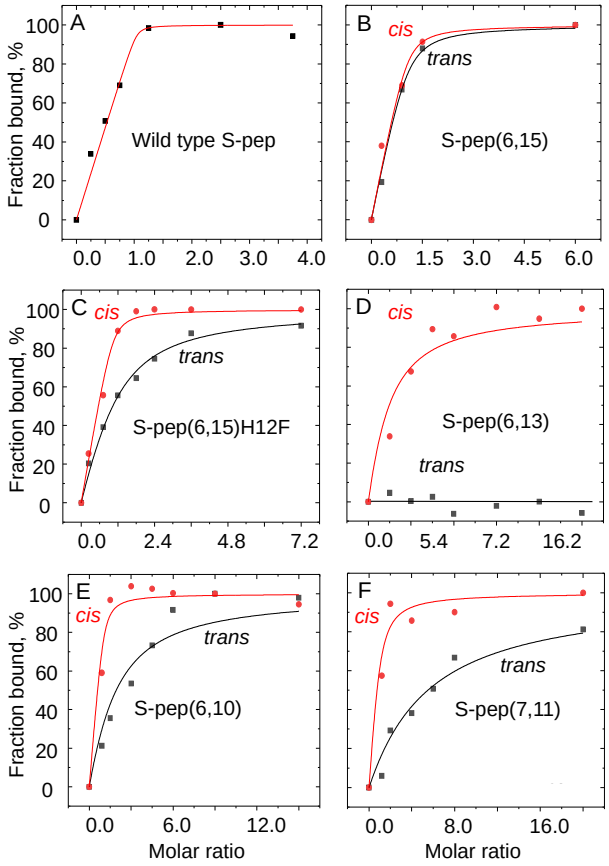


FIGURE 4.4: CD binding curves for various S-peptide variants. (A) Wild type, (B) S-pep(6,15) (C) S-pep(6,15)H12F, (D) S-pep(6,13), (E) S-pep(6,10), and (F) S-pep(7,11). Measurements with the photoswitch in *cis* configuration are plotted in red, those in the *trans* configuration in black.

chemical equilibrium, $P + L \rightleftharpoons PL$. Tab. 4.1 summarizes the results. It turns out that the method is more reliable for weak binders and as such complementary to ITC. That is, if the binding affinity becomes too large with K_d's much smaller than the concentration of the S-protein, the binding curve adopts a kink-like behaviour (see in particular the wild-type data in Fig. 4.4A) and the fit parameters become very insensitive to K_d; in essence it is then a single data point at the kink that determines K_d. Fortunately, both ITC and CD works reasonably well for S-pep(6,15), allowing us to cross-validate the two methods.

Finally, fluorescence quenching was employed as method to determine the binding dissociation constants for the *cis* and the *trans* forms of some of the peptides. The intrinsic tyrosine S-protein fluorescence decreases by circa 20% upon binding in all cases. Local structural aspects like changes of the chemical environment via hydrogen bonds, disulfide bridges formation/breaking and resonance energy transfer mechanism can affect the intrinsic tyrosine fluorescence quenching of RNase S.[272,273] The obtained binding curves are shown in Fig. 4.5, where the values for "0%-fractional-bound" and

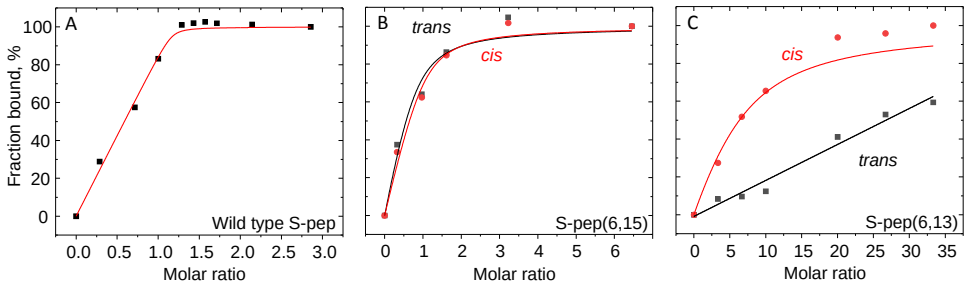


FIGURE 4.5: Fluorescence binding curves for various S-peptide variants. (A) Wild type, (B) S-pep(6,15), and (C) S-pep(6,13). Measurements with the photoswitch in *cis* configuration are plotted in red, those in the *trans* configuration in black. In the case of S-pep(6,13)-*trans*, the saturation plateau was not reached; it was assumed to be the same as for S-pep(6,13)-*cis* for the purpose of plotting (but that does not have to be correct since unspecific binding implies that more than one S-peptide might bind). The linear fit in this case is to guide the eyes.

"100%-fraction-bound" have been determined in the same way as for the CD data. Wild type and S-pep(6,15) were used to cross-validate the fluorescence method with CD and ITC techniques, with both the commercial and the home-built fluorimeters, respectively, and S-pep(6,13) was then measured as an example with very low binding affinity. The extracted binding affinities are summarized in Tab. 4.1. Overall, the results from three different methods are in good agreement with each other within experimental error.

4.4.1 Computational Results

To study the relation between changes in binding affinity (i.e, experimental K_d obtained by CD spectroscopy) and secondary structural changes of the S-peptides, all-atom MD simulations were performed as described in Methods. As an overview, Fig. 4.6 shows the resulting dynamics of secondary structures of the various S-peptides obtained from DSSP analysis.[274,275] In particular, we focus on the differences of helicity when changing from *cis* to *trans* configuration. In the case of S-pep(6,15) and S-pep(6,15)H12F, we find that the core region formed by residues 4-11 exhibits a stable α-helical conformation in both *cis*- and *trans*-configuration. Recalling that an α-helix of residues Thr3-Met13 is the hallmark of the native, active RNase S,[104] we conclude that the structure of these S-peptides are hardly perturbed by the photoswitch. This conclusion is also supported by the fact that the wild-type peptide exhibits about the same helicity as S-pep(6,15) and S-pep(6,15)H12F (see Fig. S3). S-pep(6,13), on the other hand, is significantly more affected by *cis* $\rightarrow$ *trans* photoswitching. Apart from infrequent short perturbations, S-pep(6,13)-*cis* reveals a relatively stable α-helical structure, while its *trans* form mostly exist as a 3_{10} helix or random coil. In a similar way, S-pep(6,10) and S-pep(7,11) exhibit decreased α-helicity upon *cis* $\rightarrow$ *trans* photoswitching, although the structural changes

found in *trans* are less prominent compared to S-pep(6,13). Performing a time average over the trajectories, the above findings can be represented by a probability distribution of the α-helicity of the various systems (see Fig. S3, Supporting Information).

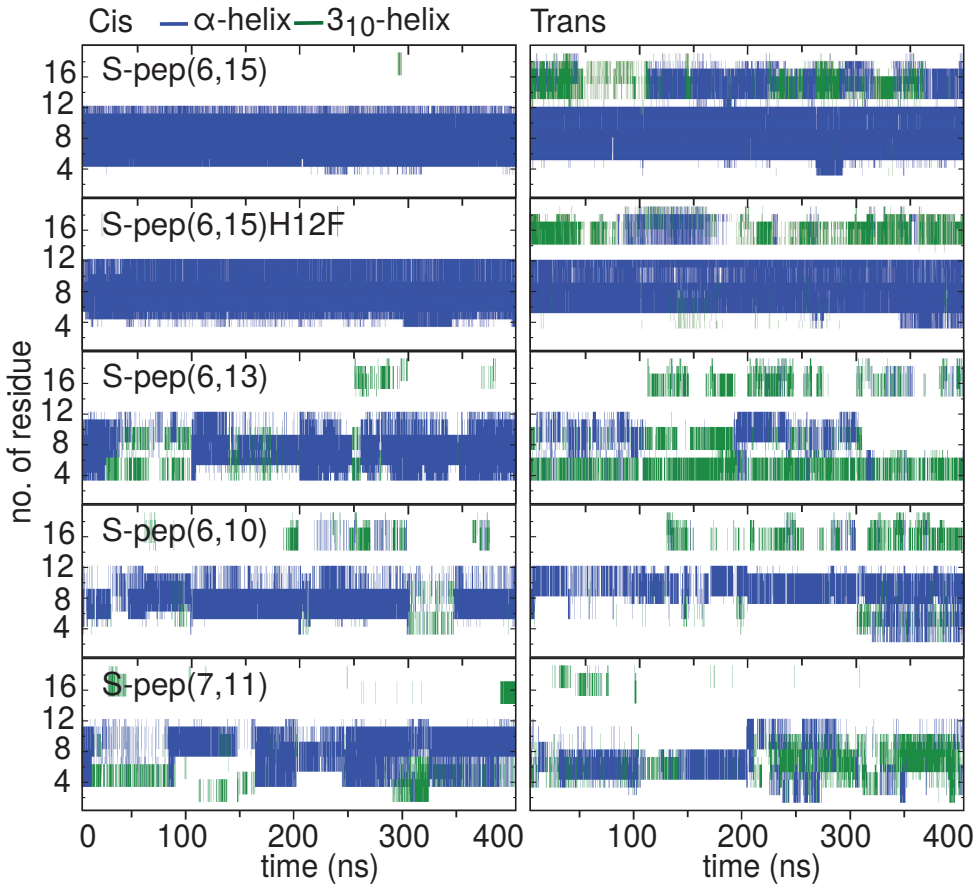


FIGURE 4.6: Time evolution of the S-peptides' secondary structure in *cis* (left) and *trans* (right) configuration, obtained from all-atom MD simulations after suitable equilibration (see Methods). α helix content in blue, 3_{10} helix in green, unstructured parts in white.

To connect these computational results for the α-helicity α_H to the experimental findings of K_d in form of a quantitative structure activity relationship (QSAR), we used a Boltzmann statistics-based ansatz, which assumes that the free energy of binding, $\Delta G \sim \ln(K_d)$, is linearly dependent on the number of peptide residues in α-helical conformation. This gives

$$\alpha_H = -c_\alpha k_B T \ln(K_d) + \alpha_{H_0}, \tag{4.1}$$

where c_α accounts for the free energy gain per helical residue $\Delta G = \alpha_H / c_\alpha$, and α_{H_0} is a reference peptide helicity. As shown in Figure 4.7, this ansatz reveals a convincing fit ($R^2 = 0.85$) for the *trans*-state. The fit is less clear ($R^2 = 0.65$) for the *cis*-state, where we also added the results of the wild-type. The resulting fit parameters for *trans* (c_α = 3.4 $\pm$ 0.9, α_{H0} = 53 $\pm$ 6) and *cis* (c_α = 4.2 $\pm$ 2.5, α_{H0} = 44 $\pm$ 6) agree with each other within their error ranges.

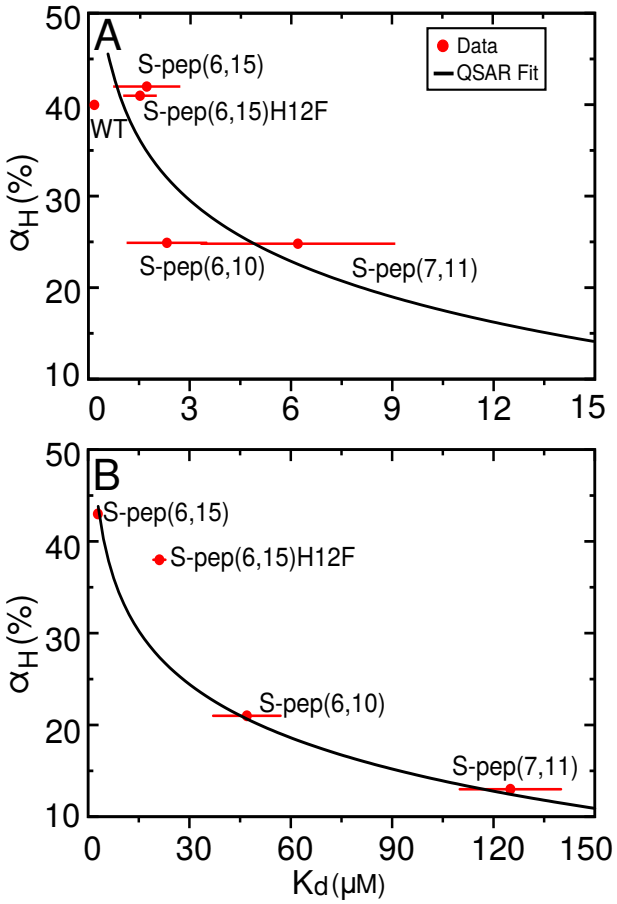


FIGURE 4.7: (A) α-helicity vs experimental K_d in *cis*-configuration, together with the wild-type (WT), and fit obtained by Eq. 4.1. (B) The same for the *trans*-configuration.

Figure 4.8 combines these results and correlates the change of the calculated change in α-helicity $\Delta\alpha_H = \alpha_{H(cis)} - \alpha_{H(trans)}$ to the experimentally measured change in affinity $\Delta K_d = K_{d(trans)} - K_{d(cis)}$. In nice agreement, both simulated $\Delta\alpha_H$ and experimental ΔK_d show an ascending behaviour starting with S-pep(6,15) via S-pep(6,15)H12F and S-pep(6,10) to S-pep(7,11) and S-pep(6,13).

It is interesting to connect the above discussed helicity changes with the prevailing molecular structures of the S-peptides. To this end, robust density-based clustering[54] (see Methods) was performed for all systems and the resulting main conformational states were analyzed in detail (see Figs. S4 and S5, Supporting Information). Overall, we find that all S-peptides can be well described by only four conformational states, where the dominant state typically occurs with 40-70% population. Figure 4.9 shows the resulting main structure of all S-peptides (the S-protein is not shown for clarity). In line with the discussion of Fig. 4.6, we find that *cis* $\rightarrow$ *trans* photoswitching has the least impact on the secondary structures of S-pep(6,15) and S-pep(6,15)H12F, while the most significant changes are found for S-pep(6,13), which essentially loses its α-helical conformation. Coloring residues according to their fluctuations, we find that the

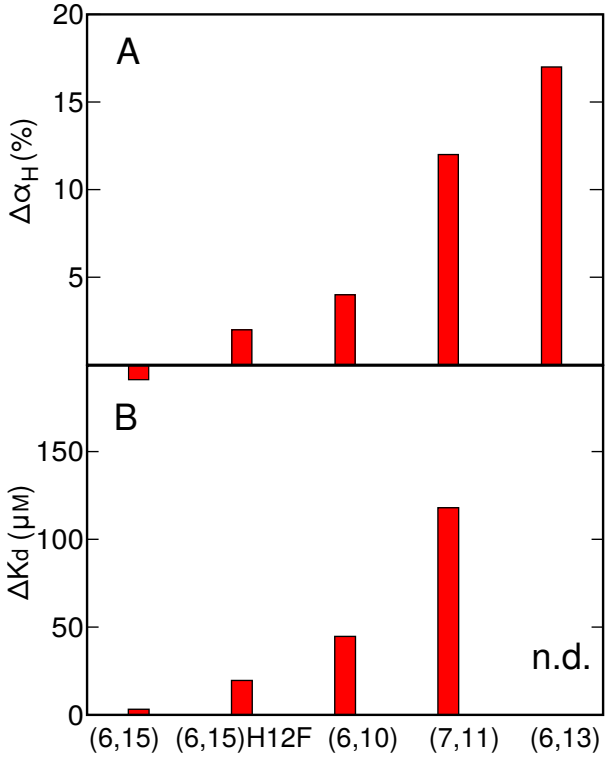


FIGURE 4.8: (A) change in α-helicity difference: $\Delta\alpha_H = \alpha_{H(cis)} - \alpha_{H(trans)}$ and (B) change in experimental $\Delta K_d = K_{d(trans)} - K_{d(cis)}$, as determined from CD spectroscopy (Table 4.1). No experimental exist for K_d in the *trans*-state of S-pep(6,13), which is why no bar is shown in panel (B), but it is clear from Figs. 4.4D and 4.5C that the difference is large.

structures are well defined in both *cis* and *trans* and exhibit increased flexibility only at the terminals. That is, even the apparently disordered regions of S-pep(6,13) and S-pep(7,11) reveal stable structures.

4.5 Discussion and Conclusion

To facilitate photocontrolling of peptide-protein binding, we have designed various photoswitchable S-peptides, switching their helical content, and measured their binding affinities to the S-protein using a variety of experimental techniques (Table 4.1). Alpha helices are structural motifs that very commonly are relevant for protein-protein, protein-peptide and protein-DNA interactions and therefore represent important targets for modulation of binding affinities. Here, we have introduced a concept by which the helical content can be modulated in a very controlled manner. The overall disruption of secondary structures not only depends upon the distance between anchor residues of photoswitch but also on the exact position of the photoswitch.

By performing MD simulations of the *cis* and *trans* configurations of the RNase S complex, we were able to relate simulated α-helicities α_H to experimentally obtained binding affinities K_d, which are in surprisingly good agreement (Fig. 4.8). Using the

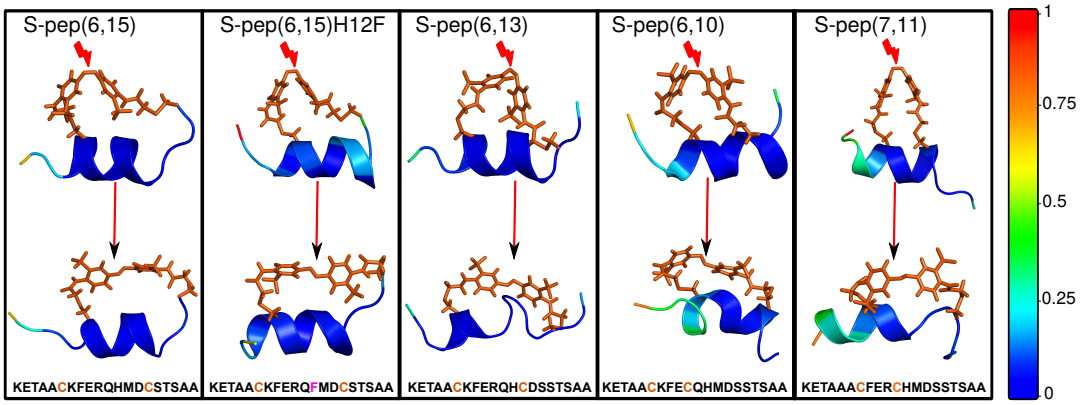


FIGURE 4.9: Most populated structures obtained from a clustering analysis, see text and Supporting Information for details. The S-protein is not shown for clarity. The residues are colored according to normalized root mean square fluctuations.

simple QSAR ansatz of Eq. 4.1, we have shown that the α-helicity is inversely related to the experimental K_d's (Figs. 4.7). This suggests that relatively short (sub-μs) MD simulations are indicative of the long-time binding or dissociation behavior of the RNase S complex. The secondary structure analysis of the S-peptides in (Fig. 4.9) shows a significant decrease in α-helicity upon *cis*-to-*trans* isomerization for all systems, with the one exception of S-pep(6,15). The latter is in line with the findings of Ref.[99] studying the photoswitching of isolated helical peptides, which revealed that the (i,i+9) spacing in S-pep(6,15) is the dividing point between decreasing or increasing the α-helical content upon *cis*-to-*trans* isomerization.

The largest change in α-helicity is found for S-pep(6,13) and S-pep(7,11) with spacings (i,i+7) and (i,i+4), respectively. The corresponding structures of the *trans* forms in Fig. 4.9 reveal a considerable loss of α-helical conformation. In particular, S-pep(6,13)-*trans* adopts a helical conformation only at its N-terminal, while its C-terminus is completely disordered. It has been suggested that the N-terminal part of the S-peptide is unzipping from the S-protein binding groove due to equilibrium state fluctuations.[106] Furthermore, based on an alanine mutant screening approach,[106] it has been shown that the N-terminus (more precisely residues 1-7) contributes to binding only to a minor extent. Hence, even though a small percentage of helicity was preserved in the case of S-pep(6,13)-*trans*, no specific binding is observed, since the residual helicity is located in relatively unimportant part for the binding.

It is interesting to note that the isomerization of the photoswitch does not lead to a significant increase of peptide fluctuations. Instead, Figs. 4.9 and S4 (see Supporting Information) suggest that peptide structures are mostly well defined in both the *cis* and the *trans*-state and exhibit fluctuations only at the termini. Especially the parts of the peptide that lie between the clamping points of the photoswitch display highly stable structures, even in apparently disordered states. Furthermore, in the case of S-pep(6,13),

the peptide structure is more stable in the disordered *trans* state than in helical *cis* state, in clear contrast to intuition. This stabilization of disordered states turns out to be one of the most interesting outcomes of our photolabeling strategy. In combination with structural predictions from MD simulations, it allows a rather distinct manipulation of the peptide structure depending on the photoswitch attachment points.

We consider S-pep(6,13) the most interesting result in our series of mutants. Firstly, while no binding could be detected for S-pep(6,13)-*trans* via CD spectroscopy (Fig. 4.4D), some degree of binding is observed by fluorescence quenching (Fig. 4.5C), illustrating the different aspects of binding the two methods are sensitive to. That is, while CD measures helical content, fluorescence quenching measures spatial proximity. The difference between CD and fluorescence data can probably be explained by the fact that binding is unspecific in this case, i.e., more in the sense of aggregation, and hence does not induce any α-helical structure. Since we did not reach a plateau in fluorescence intensity in this case, we cannot truly quantify the binding affinity, but it is probably in the mM range. And indeed, such weak binding can typically no longer be characterized as "specific", in particular if the binding partner is a floppy peptide like in the present case.[276]

Secondly, the largest effects observed in literature for photoswitchable protein-peptide complexes are K_d ratios of about 20 (Ref.[259]), and the effect is equally large for S-pep(6,10) and S-pep(7,11). For S-pep(6,13), the ratio is even larger, i.e., effectively infinite, as no specific binding is detectable by CD spectroscopy in the *trans*-state. To the very best of our knowledge, such a big effect has not been achieved so far for any photoswitchable protein-peptide complex. This property makes the system particularly interesting for time-resolved studies, in which the light-driven isomerization of the photoswitch triggers unbinding of the S-peptide, and the structural response of the protein is investigated by e.g. transient IR spectroscopy or transient X-ray scattering experiments in upcoming free electron lasers.[239] Ideally, in such an experiment, the S-peptide should unbind as quickly as possible; only then the temporal ordering of unbinding event vs structural response could be measured, addressing the long-standing question whether the process can be described as "conformational selection" or as "induced fit". However, the unbinding rate constant k_{off} scales essentially as K_d (since k_{on} is limited by diffusion and as such essentially a constant), hence to speed up k_{off}, K_d in one state of the photoswitch should be as large as possible. Transient IR experiments are currently underway.

4.6 Supplementary Information

ITC of S-peptides in the *cis*-state

For the S-peptides in the *cis*-state, there has been a concern whether the binding affinity can be determined by ITC at all. That is, after injection of the photoswitchable peptide solution into the closed and dark sample cell, illumination was no longer possible and thermal back-isomerization of the photoswitch from *cis*-to-*trans* occurred to a certain extent. First, the time to complete the ITC measurement was kept as short as possible (45 min), which is in the same range as thermal back relaxation of the free S-peptide (between 40-60 min depending on S-peptide variant, see Fig. S1, red). We however found that thermal back relaxation is significantly longer once an S-peptide is bound to the S-protein (see Fig. S1, blue), so the time that really counts is k_{on}, i.e. the time to bind right after an injection, which is short enough to be unproblematic. Independent of that

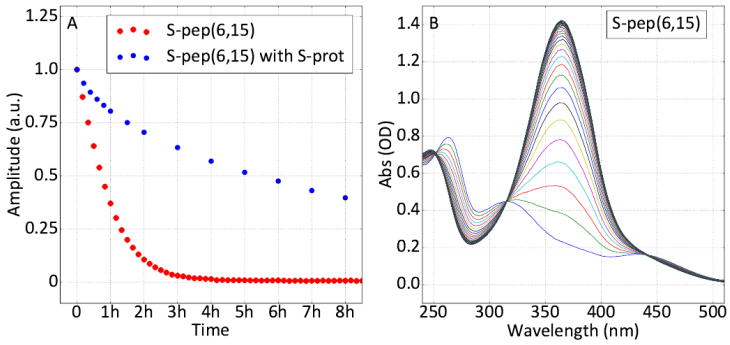


FIGURE S1: Thermal *cis*-to-*trans* back-relaxation exemplified for S-pep(6,15); in red for the isolated peptide in solution, and in blue for S-pep(6,15) bound to the S-protein. The inset shows corresponding UV/VIS absorption spectra, from which the kinetic data are retrieved, with the *trans*-band at 367 nm being reestablished upon the thermal back-relaxation.

issue, thermal back-isomerization releases significant amount of heat, and one might expect a large background signal in the ITC data. In fact the heat released by back-isomerization[264] is about the same as the heat of binding. However, back-isomerization occurs continuously, in contrast to the binding-associated heat that occurs abruptly following every injection, and high-pass filtering in the ITC instrument can suppress that background completely. As a result, the binding affinities could be measured reliably also in the *cis*-state of the S-peptide, as our cross-validation with other methods has shown (see main text).

ESTIMATE OF THE AMOUNT OF CIS -CONFIGURATION

The cis-configuration is prepared by illuminating the sample with a 370 nm cw diode laser (CrystaLaser). Provided that the power of the light source is high enough to

outperform thermal cis-trans back reaction (a condition which is safely fullfilled with that light source), this leads to a relatively high content of cis-configuration, since the trans-configuration has a strong band at this frequency position, while the absorption cross section of the cis-configuration is small. Fig. S1B (blue) shows an absorption spectrum of the cis-configuration prepared in that way, which then relaxes back to the trans-configuration. If we assume that the absorption cross section of the cis-configuration is exactly zero, we can estimate from the ratio of absorbances at 370 nm that the sample contains 85% cis-configuration. Since the absorption cross section of the cis-configuration is probably not quite zero, that estimate is considered a worst case scenario, i.e., a lower limit.

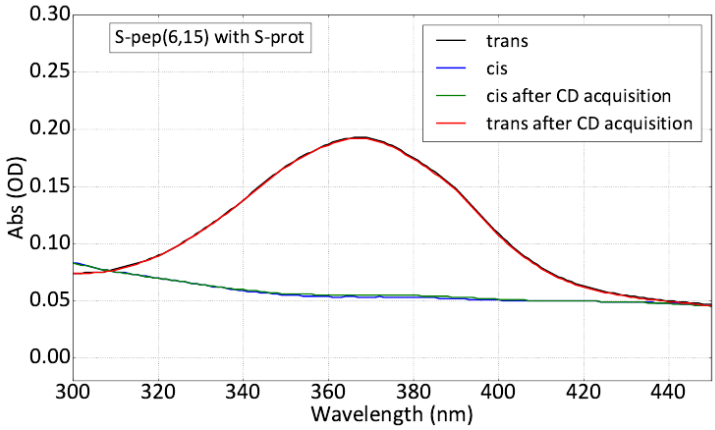


FIGURE S2: UV absorption spectra of S-pep(6,15) in the trans-configuration before (black) and after (red) measuring the actual CD data that led to Fig. 4B, and the same for the cis-configuration in blue and green, respectively

Computational Details

All systems were solvated with ca. 7000-8000 TIP4P-2005[267] water molecules in a dodecahedron box with an image distance of 7-8 nm. Na^+ and Cl^- were added at a salt concentration of 0.1 M with an excess of Cl^- to compensate the net positive charge of the complexes. Force field parameters for the azobenzene switch were taken from Ref..[159] All bonds involving hydrogen atoms were constrained using the LINCS algorithm,[206] allowing for 2 fs time step. Long-range electrostatic interactions were computed by the Particle Mesh Ewald method,[145] whereas the short-range electrostatic interactions were treated explicitly with a Verlet cutoff scheme. The minimum cutoff distance for electrostatic and van der Waals interactions was set to 1.4 nm. A temperature of 300 K was maintained via the Bussi thermostat (aka velocity-rescale algorithm)[146] with a coupling time constant of 0.1 ps. A pressure of 1 bar was controlled using the pressure coupling method of Berendsen[147] with a coupling time constant of 0.1 ps. After energy minimization, all systems were equilibrated for 10 ns using an NPT ensemble.

S-pep(6,15) was then simulated for 1 μs in both *cis* and *trans* configuration of photoswitch, S-pep(6,15)H12F for 800 ns in *cis* and 2 μs in *trans* configuration. S-pep(6,13), S-pep(6,10) and S-pep(7,11) were simulated for 400 ns in both configurations. Principal

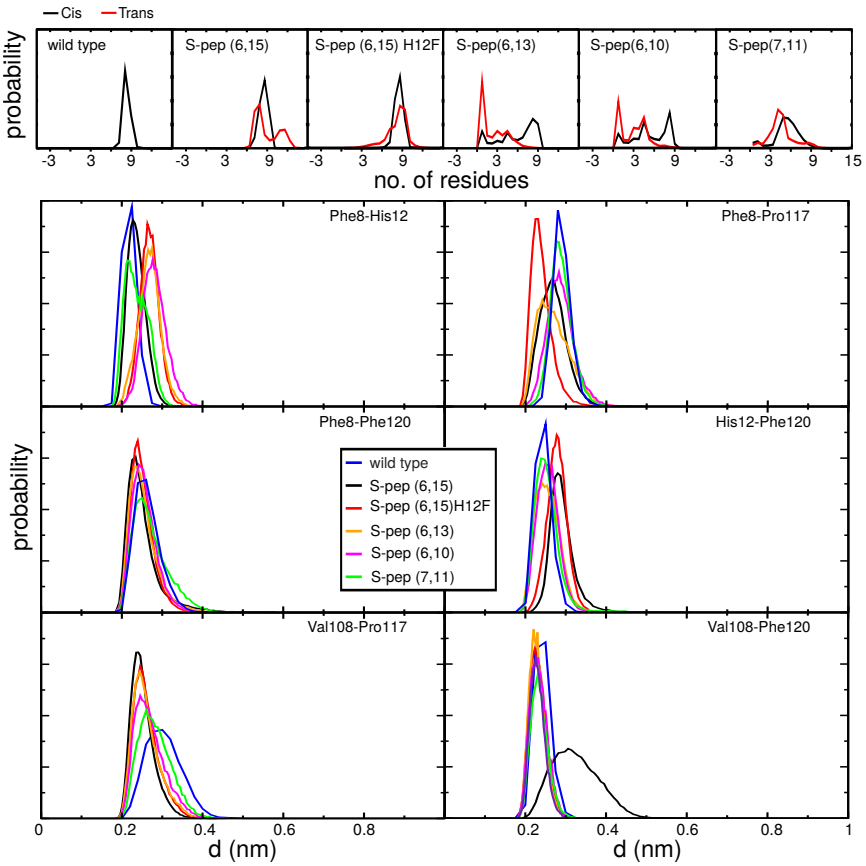
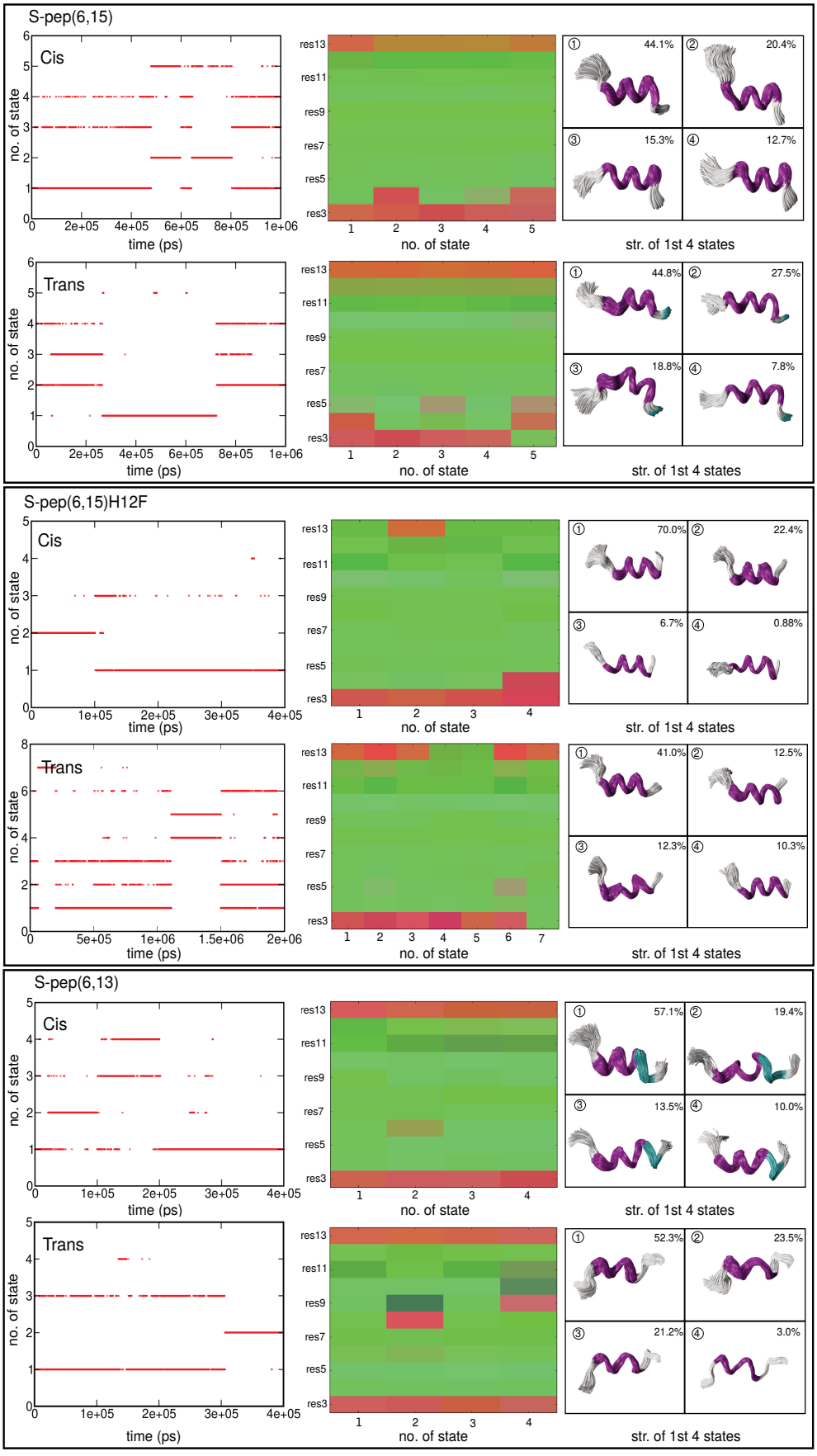


FIGURE S3: Probabilities of No. of residues in a helical conformation; in black for the *cis*-state and in red for the *trans*-state.

component analysis on backbone dihedral angles (dPCA+)[268] was carried out by diagonalizing the covariance matrix $\sigma_{ij} = \langle (\varphi_i - \langle \varphi_i \rangle)(\varphi_j - \langle \varphi_j \rangle) \rangle$ of input coordinates $\{\varphi_i\}$. Projecting the MD trajectories onto the resulting eigenvectors $v^{(i)}$ we define the principal components, $V_i = v^i \cdot (\varphi - \langle \varphi_i \rangle)$, which are naturally ordered in descending variance of the system. After dimensionality reduction, robust density-based clustering[54] was performed by computing a local free energy estimate for every structure of the trajectory via counting all other structures inside a d-dimensional hypersphere of fixed radius R, where R equaled the lumping radius.[55] Normalization of these population count yields densities, which give the free energy estimate $F_R = -k_B T \ln(P_R / P_R^{max})$ starting at $\min(F_R) = 0$.


S-pep(6,15)
Cis
Trans
no. of state
time (ps)
res13
res11
res9
res7
res5
res3
no. of state
str. of 1st 4 states
① 44.1%
② 20.4%
③ 15.3%
④ 12.7%
① 44.8%
② 27.5%
③ 18.8%
④ 7.8%
S-pep(6,15)H12F
① 70.0%
② 22.4%
③ 6.7%
④ 0.88%
① 41.0%
② 12.5%
③ 12.3%
④ 10.3%
S-pep(6,13)
① 57.1%
② 19.4%
③ 13.5%
④ 10.0%
① 52.3%
② 23.5%
③ 21.2%
④ 3.0%

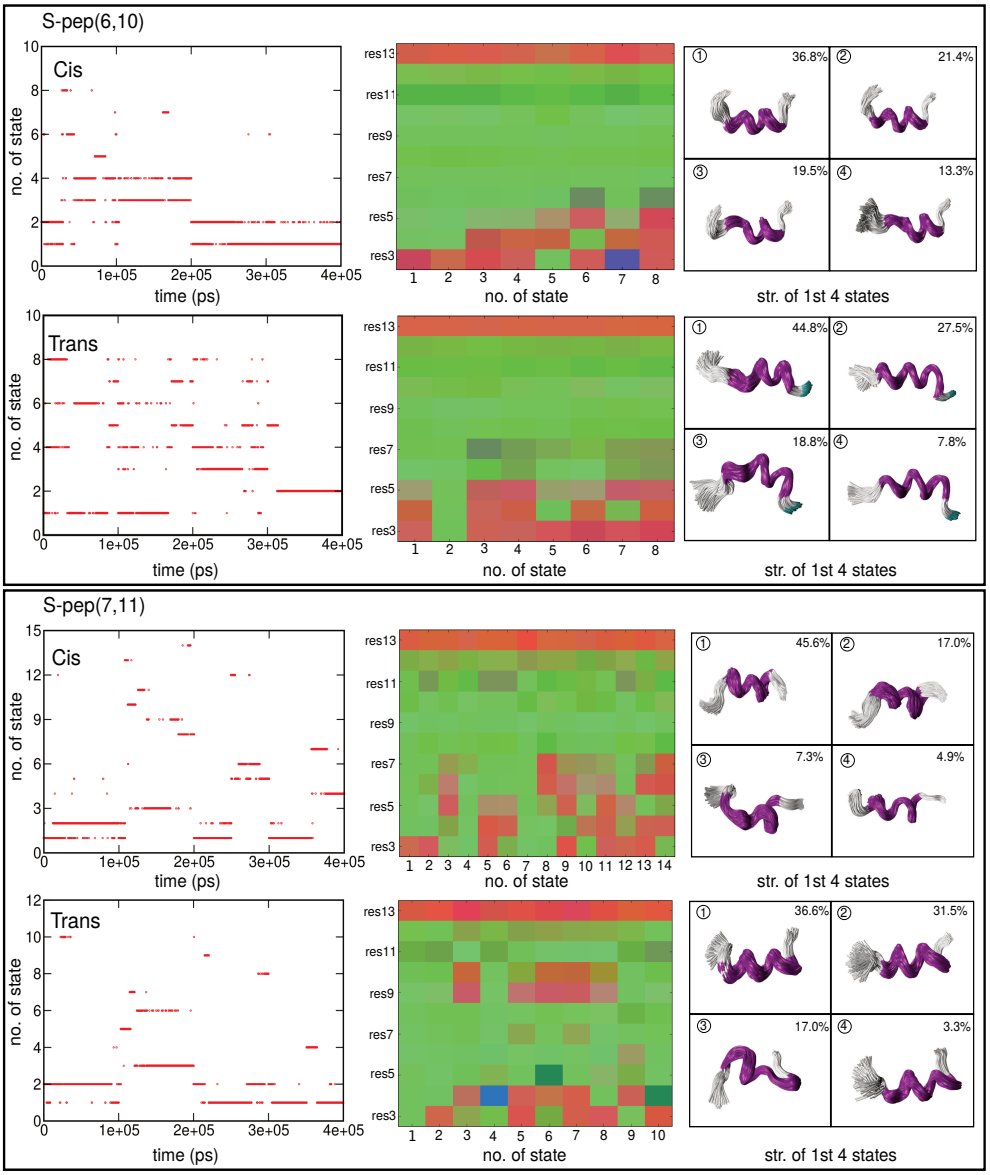


FIGURE S4: State trajectory (No. of states vs No. of frames), ramacolor plot and representative structures of first four most populated states of all the considered systems in both *cis* and *trans*. The colors of the ramacolor plots are specified in Fig. S5.

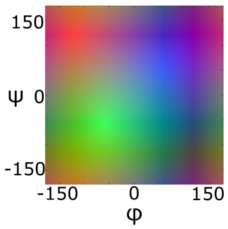


FIGURE S5: Color space spanned by ϕ- and ψ-coordinates define as: red is associated with extended conformations (β, P_{II}) centered around (-120, 120), green with right-handed helical conformations (α_R, 3_{10R} centered around (-60,-60), and blue with left-handed helical conformations (α_L, 3_{10L}) centered around (60, 60).

Chapter 5

Signal Transport in Biomolecules

This chapter is based on following publication:
Real-time observation of ligand-induced allosteric transitions in a PDZ domain
Olga Bozovic[a,1], Claudio Zanobini[a,1], Adnan Gulzar[b,1], Brankica Jankovic[a], David Buhrke[a], Matthias Post[b], Steffen Wolf[b], Gerhard Stock[b,] and Peter Hamm[a,*] Submitted*

Department of Chemistry, University of Zurich, Zurich, Switzerland
Biomolecular Dynamics, Institute of Physics, Albert Ludwigs University, Freiburg, Germany
[1]*authors contributed equally*
[*]*corresponding authors: peter.hamm@chem.uzh.ch and stock@physik.uni-freiburg.de*

In a joint experimental-theoretical approach which includes protein engineering, photoswitching and time-resolved vibrational spectroscopy, as well as non-equilibrium molecular dynamics simulations with subsequent dimensionality reduction and Markov modeling, we characterize the ligand-induced conformational transitions in a PDZ2 domain, which are thought to be responsible for the allosteric communication within the protein. To prepare a non-equilibrium condition, an azobenzene-derived photoswitch is linked to a peptide ligand in a way that its binding affinity to the PDZ2 domain changes upon switching. The PDZ2 domain, in contrast, is kept unmodified, hence ligand-switching mimics very closely the naturally occurring allosteric perturbation caused by ligand (un)binding events. The response of the PDZ2 domain is best described as a remodelling of its rugged free energy landscape, with subtle but notable shifts in the population of a small number of structurally well defined states. The dynamics of the allosteric transition cover four decades of time, ranging from $\sim$1 ns to $\sim$10 μs, with a discrete set of time-scales. It is proposed that no clear-cut separation between purely dynamically driven allostery and allostery upon a conformational change exists.
A preliminary analysis of α_3-switched PDZ3 domain reveals that the direct change in contact formation betwenn ligand and α_3 caused a change in binding affinity of ligand to PDZ3 domain by a 25-folds.

Real-time observation of ligand-induced allosteric transitions in a PDZ domain
Here, I performed all MD simulations and analyzed the simulation data.

5.1 Ligand-induced allosteric transitions in a PDZ domain

Allostery is the coupling of two sites in a protein or a protein complex, where the binding of a ligand to the distal site modifies the affinity at the active site.[18] Since biological function is intimately related to protein structure, ligand-induced changes of the protein's function (e.g., the transition from an inactive to an active state) are often associated with a change of the protein's mean structure.[110] On the other hand, ligand (un)binding may also alter the protein's flexibility, which changes the variance of the structure and gives an entropic contribution to the free energy.[26] Referring to the associated change of the structural fluctuations, the latter scenario, termed "dynamic allostery," has been invoked to explain apparent absence of conformational change upon ligand (un)binding.[24,26,27,112,113,277–279] Studying the effects of dynamic allostery has been mainly done by NMR spectroscopy[280–282] which, however, only accounts for equilibrium dynamics.

While both models, structural change vs. dynamic change, may appear plausible, a stringent verification ultimately requires us to study the genesis of allostery. This includes three steps: (1) The (un)binding of a ligand (usually initiated by a change of its concentration[283]) causes (2) the atoms of the protein to undergo a non-equilibrium time evolution, which (3) eventually leads to a change of the binding affinity at a remote site of the protein. This so-called "allosteric transition" is a non-equilibrium process and has been observed directly only rarely, in part because the smallness of the structural changes makes the transition pathways challenging to observe experimentally,[284] and also because of the time-scale limitations of molecular dynamics (MD) simulations.[285–287] In this work, we outline an approach to study the first two steps, i.e., the ligand-induced allosteric transition, employing a PDZ2 domain as model system.

Known for their modest conformational change upon ligand binding, PDZ domains are considered as prime examples of dynamic allostery.[111–113] PDZ domain-mediated interactions play a pivotal role in many signal transduction complexes.[288,289] Allosteric information flow in PDZ domains is thought to be transduced via conserved allosteric networks in the protein.[30,112,128,290,291] The system considered here is the PDZ2 domain from hPTP1E (human tyrosine phosphatase 1E) and a RA-GEF-2 peptide derivative (Ras/Rap1 associating guanidine nucleotide exchange factor 2)[114] with an azobenzene moiety linked as photoswitch,[115] see Fig. 5.1. It was recently reported for a very similar system that the phosphorylation of the serine (-2) residue, a common target in regulatory processes of PDZ domains,[292] leads to a $\sim 5-7$-fold difference in the affinity

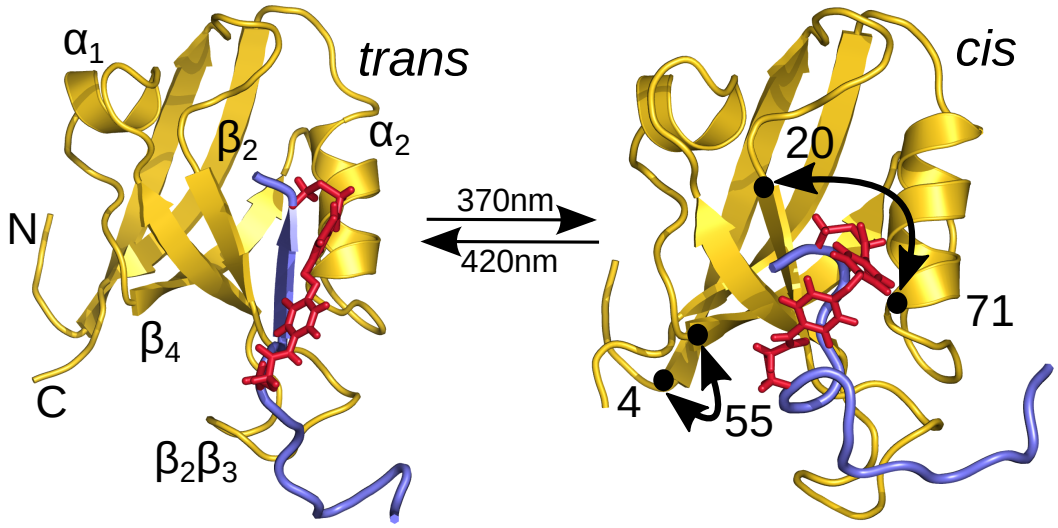


FIGURE 5.1: Ligand-switched PDZ2 domain. Main secondary structural elements and C_α-distances $d_{20,71}$ and $d_{4,55}$ discussed below are indicated. In the *trans* conformation of the photoswitch (red), the ligand (blue) fits well in the binding pocket, while it starts to move out when switching to *cis*.

towards the PDZ2 domain.[293] We will see that the binding affinity can be perturbed to the same extent (≈5 fold) by introducing such a photoswitchable element on the ligand instead. Since the PDZ2 domain is not modified at all, this strategy leads to a much less artificial construct than obtained in our previous study,[16] where the photoswitch was covalently linkedacross the binding pocket of the PDZ2 domain.

By photo-isomerizing the azobenzene moiety, we change the binding affinity of the ligand at a precisely defined point in time. We employ isotope-labelled time-resolved vibrational spectroscopy to monitor the structural change of the protein in real time, and perform extensive (more than 0.5 ms aggregate simulation time) all-atom non-equilibrium MD simulations combined with Markov modeling to interpret the experimental results in terms of the structural evolution of the system. We find that the mean structural change of the protein is rather small. Yet, in both experiment and MD simulations the free energy surface of the protein can be characterized by a small number of metastable conformational states. In agreement with the view of allostery as an interconversion between the relative population of metastable states, we see how the ligand-induced response of the PDZ2 domain is best described as remodelling of the free energy landscape,[116–120] and how the response is transduced from the ligand to the protein without introducing a significant structural change.

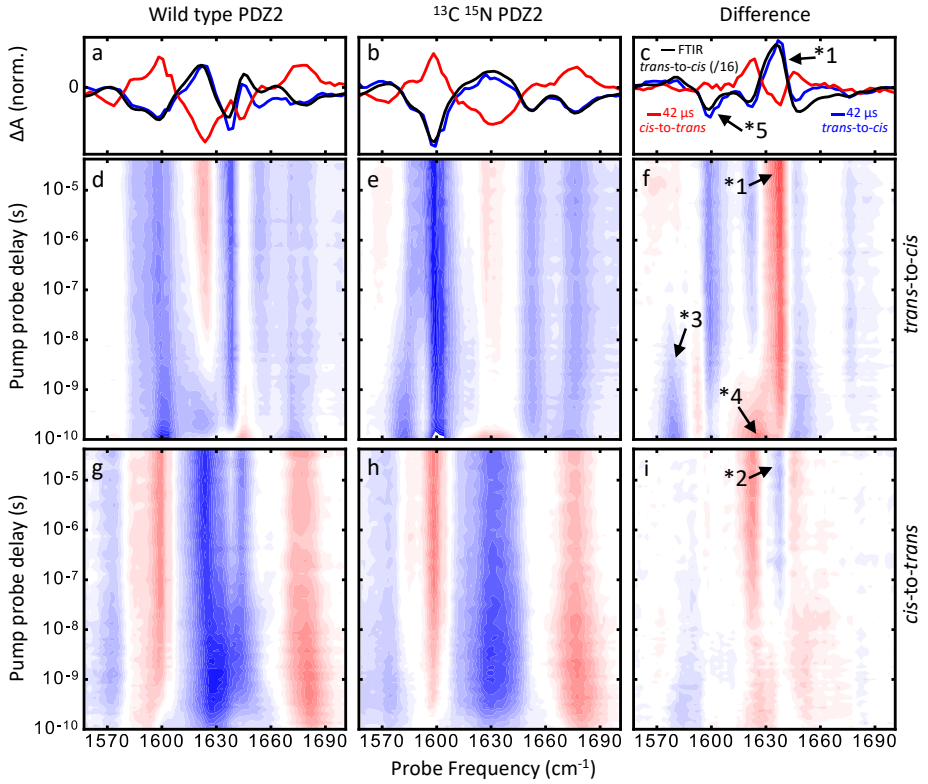


FIGURE 5.2: Transient IR spectra of PDZ2 in the region of the amide I band. Panels (a-c) compare transient data at long pump-probe delay times (averaged from 20 μs to 42 μs to increase signal-to-noise) for *trans*-to-*cis* (blue) and *cis*-to-*trans* (red) switching, together with a properly scaled *trans*-minus-*cis* FTIR difference spectrum (black). Panels (d-f) show the complete transient data for *trans*-to-*cis* switching, and panels (g-i) for *cis*-to-*trans* switching. Left panels show the data for the wild type (WT) protein, middle panels for the sample with the protein $^{13}C^{15}N$ labelled (the peptide ligand contains naturally abundant $^{12}C^{14}N$), and right panels the $^{13}C^{15}N$-WT difference data. Red colours in panels (d-i) indicate positive absorbance changes, blue colors negative absorbance changes. The relative scaling of the data sets and the labelled features are discussed in the text.

5.2 Results

5.2.1 Experimental

To set the stage, we have investigated the influence of photoswitching of the ligand on its binding affinity. By choosing the spacing between the anchoring points of the azobenzene moiety, the peptide ligand was designed such that the longer *trans* conformation mimics the native extended β-strand conformation, while the *cis* configuration shortens the peptide and perturbs it from its extended form. To that end, the alanine residue at position -1 (the ligand is labelled by negative numbers) was chosen as the first anchoring spot for the photoswitch, since it has been shown that a mutation at this

position does not significantly affect the binding, while residues that are crucial for binding (Val(0), Ser(-2) and Val(-3)) are preserved.[127,294] The second anchoring point chosen was Asp(-6) which allows the peptide to be maximally stretched in the *trans* configuration of the photoswitch. Protein and peptide have been expressed/synthesized using standard procedures,[16,159] see Supplementary Methods and Supplementary Fig. S1 for details. The dissociation constants (K_D) in the two configurations of the photoswitchable peptide were determined by ITC, fluorescence and CD spectroscopy (see Supplementary Figs. S2 and S3).[295] The obtained values averaged for all methods ($K_{D,trans} = 2.0 \pm 0.6$ μM, $K_{D,cis} = 9.6 \pm 0.5$ μM, see Supplementary Table 1) reveal an appreciable $\sim$5-fold difference in the binding affinity, with the *cis* state being the destabilized one, as anticipated.

Considering these binding affinities and the relatively high concentrations needed for the transient IR experiment (1.25 mM for the peptide and 1.5 mM for the protein), it is clear that most of the ligands are bound in both states to a protein of the photoswitch (97% in *cis* and 99% in *trans*), hence we will not observe many binding or unbinding events. Furthermore, as binding and unbinding in similar PDZ/ligand systems was observed to occur on 10 – 100 ms time-scales,[190] these processes are hardly within the time window of our experiment. Nevertheless, we will be able to observe the adaptation of the protein to a perturbed peptide conformation in the binding pocket and its transition to unspecific binding on the protein surface.

We investigate that process with the help of transient IR spectroscopy in the range of the amide I band (see Supplementary Methods for details).[198,296,297] This band originates from mostly the C=O stretch vibration of the peptide/protein backbone, and is known to be strongly structure dependent.[298] While one cannot invert the problem and determine the structure of a protein from the amide I band, any change in protein structure will cause small but distinct changes in this band (see Fig. 5.2 a-c).

Figure 5.2 shows the transient IR response in the spectral region of the amide I vibration after photoswitching in either the *trans*-to-*cis* (panels d-f) or the *cis*-to-*trans* direction (panels g-i). To be directly comparable, the two data sets were scaled in a way that they refer to the same amount of isomerizing molecules, and not the same amount of excited molecules. The scaling took into account the different pump-pulse energies used in the experiments (see Supplementary Methods), cross sections (23500 $cm^{-1}M^{-1}$ for *trans* at 380 nm vs 2000 $cm^{-1}M^{-1}$ for *cis* at 420 nm),[115] and isomerization quantum yields (8% for *trans*-to-*cis* switching and 62% for *cis*-to-*trans* switching).[299]

The left panels of Fig. 5.2 show the results for the wild type PDZ2 domain, and the middle panels those with the protein $^{13}C^{15}N$ labelled, which down-shifts the frequency of the amide I band by $\approx$25 cm^{-1}. The transient IR responses of both isotopologues look quite similar, as the signal is dominated by the photoswitchable peptide, which is perturbed directly by the azobenzene moiety. To remove that contribution and to isolate the smaller protein response, the two signals have been subtracted in the right

panels of Fig. 5.2, with some of the more prominent features highlighted in Fig. 5.3a-d. In this way, we take advantage of the fact that only the amide I band of the protein is affected by $^{13}C^{15}N$-isotope labelling and not that of the photoswitchable ligand. By doing so, we implicitly assume that the spectra of protein and ligand are additive and that coupling between them can be neglected. Great care was taken that protein and peptide concentrations were exactly the same in both experiments. Furthermore, both experiments were performed right after each other without changing any parameter of the laser setup.

Overall, the kinetics of these double-difference spectra are quite complex and cover many orders of magnitudes in time.[300] Furthermore, the responses for *trans*-to-*cis* (Figs. 5.2f and 5.3a,c) vs *cis*-to-*trans* switching (Figs. 5.2i and Fig. 5.3b,d) are not mirror-images from each other, which one might expect if the protein would take the same pathway in the opposite direction. For example, the strongest band at 1636 cm^{-1} (marked as *1 in Figs. 5.2f and 5.3a) reveals the biggest step at around 1 ns in the *trans*-to-*cis* data, while the complementary feature in *cis*-to-*trans* data (marked as *2 in Figs. 5.2i and 5.3b) develops in a very stretched manner from ≈3 ns to ≈3 μs. Worthwhile noting is also a transient band at 1579 cm^{-1} in the *trans*-to-*cis* data (marked as *3 in in Figs. 5.2f and 5.3c), living up to ≈100 ns, which has no complementary counterpart in the *cis*-to-*trans* data (Figs. 5.2i and 5.3d).

The red lines in Figs. 5.3a-d are fits revealed from a time-scale analysis of the signals using a Maximum Entropy Methods (see Supplementary Methods for details):[301–303]

$$S(\omega_i, t) = a_0(\omega_i) - \sum_k a(\omega_i, \tau_k) e^{-t/\tau_k}. \tag{5.1}$$

Here ω_i denotes the probe frequency and t the delay time of the signal, which is represented by a multiexponential function with time-scales τ_k. The time-scale spectra $a(\omega_i, \tau_k)$ are shown in Figs. 5.3a-d as blue lines. Each of the kinetic processes discussed above shows up as a peak in these time-scale spectra, and the pattern of peaks is different for all the examples shown in Figs. 5.3a-d. Nevertheless, the dynamical content,[14] $D(\tau_k) = [\sum_i a^2(\omega_i, \tau_k)/n]^{1/2}$, which averages over the complete data set shown in Supplementary Fig. S6, seems to indicate a relatively small number of discrete time scales, see Figs. 5.3e,f. We attribute the first peak around 100 ps (labelled as *4 in Figs. 5.2f and 5.3e) to a "heat signal" originating from the vibrational energy released by the photoisomerization of the azobenzene moiety, an effect that is seen universally in this type of experiments.[87,304]

The transient spectra at the latest pump-probe delay time that is accessible to our transient experiment (i.e., 42 μs) are shown in Figs. 5.2a-c in blue for *trans*-to-*cis* switching and in red for *cis*-to-*trans* switching. They are compared to a properly scaled *trans*-minus-*cis* FTIR difference spectrum (black), which represents the response at effectively infinite time after photoswitching. The counterpart of the negative band in the blue and

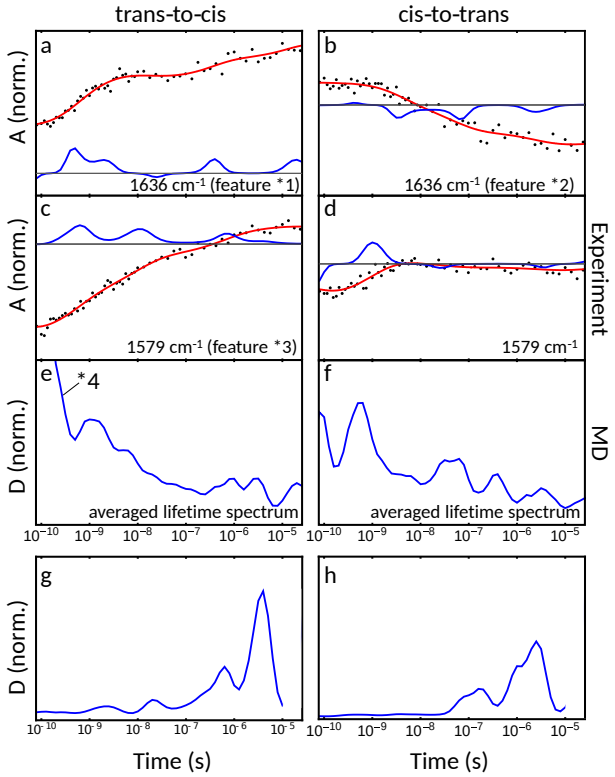


FIGURE 5.3: Transient $^{13}C^{15}N$-WT difference data at 1636 cm^{-1} (panels a,b) and 1579 cm^{-1} (panels c,d) for *trans*-to-*cis* (left) and *cis*-to-*trans* (right) switching, highlighting features labelled as *1 to *3 in Fig. 5.2. Red lines are fits obtained from the time-scale analysis in Eq. (5.1), blue lines represent the resulting time-scale spectra $a(\omega_i, \tau_j)$. Panels (e,f) show the corresponding dynamical content; the heat signal labelled as *4 is discussed in the text. Panels (g,h) show the MD dynamical content, obtained from a time-scale analysis of the non-equilibrium time evolution of the mean C_α-distances (Supplementary Fig. S7).

black *trans*-to-*cis* spectra at 1600 cm^{-1} (marked as *5 in Figs. 5.2c) has not yet evolved in the red *cis*-to-*trans* spectrum. We conclude from this observation that the *cis*-to-*trans* transition is not completely finished after 42 μs.

5.2.2 MD simulations

To aid the interpretation of the above experiments, we performed all-atom explicit-solvent MD simulations of the *cis* and *trans* equilibrium states as well as non-equilibrium MD simulations[40] of the ligand-induced conformational changes of PDZ2. Using the GROMACS v2016 software package[48] and the Amber99*ILDN force field,[139–141] we collected in total 510 μs simulation time (see Supplementary Methods). For the structural characterization of the protein, we determined 56 C_α-distances $d_{i,j}$ between residues i and j that are not redundant (such as $d_{i,j}$ and $d_{i,j\pm1}$) and whose ensemble average

changes significantly ($\langle \Delta d_{ij} \rangle \geq 0.5$ Å) during the non-equilibrium simulations (Supplementary Fig. S7). To identify the essential coordinates of the system, we performed a principal component analysis on the normalized distances of all simulation data,[53] followed by robust density-based clustering[54] and a recently proposed machine learning approach[305] (see Supplementary Methods and Fig. S8 for details). While we used six dimensions for the clustering, we find that two C_α-distances suffice to qualitatively characterize the conformational distribution of PDZ2: $d_{20,71}$ accounting for the width of the binding pocket located between β_2 and α_2, as well as $d_{4,55}$ representing the distance between N-terminus and α_1-β_4 loop, which reflects the compactness of the C- and N-terminus region (see Fig. 5.1). Employing these coordinates, Fig. 5.4a shows the free energy surface $\Delta G = -k_\mathrm{B}T \ln P(d_{20,71}, d_{4,55})$, obtained from $5 \times 5\mu$s-long *trans* equilibrium simulations describing the ligand-bound state of PDZ2. The free energy landscape reveals four well-defined local minima indicating metastable conformational states of the system. Density-based clustering identifies state **1** as close to the crystal structure,[188] while state **2** indicates an opening of the binding pocket. Both states are mirrored by states **3** and **4**, which are shifted to larger values of coordinate $d_{4,55}$.

Upon switching the ligand from *trans* to *cis* configuration, PDZ2 undergoes a non-equilibrium time evolution until it relaxes within a few microseconds (see below) into its *cis* equilibrium state, describing the perturbed protein-ligand complex. Performing $25 \times 10\,\mu$s-long *trans*-to-*cis* non-equilibrium simulations, we took the last 7μs of each trajectory to estimate the rather heterogeneous conformational distribution of the *cis* equilibrium state. When we compare the resulting free energy landscapes of *cis* and *trans*, Figs. 5.4a,b reveal that the accessible conformational space in *cis* is considerably increased, along with the occurrence of additional state **5** that reports on a further opening of the binding pocket. Representing the populations of all states in *trans* and *cis* as a histogram, Fig. 5.4d demonstrates that the photoswitching of the ligand causes a notable ($\lesssim 20\,\%$) shift of the state populations, mostly from state **1** to states **2** and **5**.

To illustrate the conformational changes associated with these states, Fig. 5.4e displays an overlay of minimum-energy structures of states **1** and **2** as well as the *cis*-specific state **5**. We find that the opening of the binding pocket described by $d_{20,71}$ mainly reflects a shift of the α_2 helix down and away from the protein core. Interestingly, the structural rearrangement between main states **1** and **2** results in an overall root mean squared (RMS) displacement of only $\lesssim 1$ Å and causes only few (~ 5) contacts to change (Supplementary Fig. S9). This is in striking contrast to the cross-linked photoswitchable PDZ2 studied by Buchli et al.[16] where 34 contact changes were found for the *trans*-to-*cis* reaction,[186] and *cis* and *trans* free energy landscapes hardly overlapped.[14] This findings indicate that ligand-switching is considerably less invasive than a cross-linked photoswitch and therefore better mimics the natural unbiased system.

Is the above discussed population shift as well as the very occurrence of states an

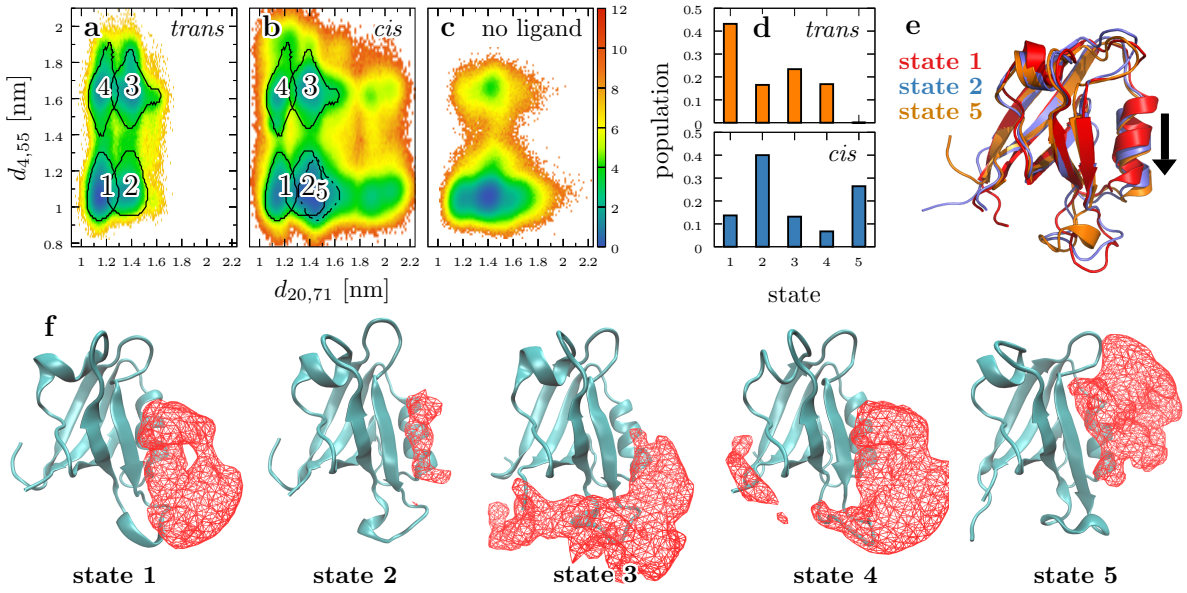


FIGURE 5.4: Identification of metastable conformational states. Free energy landscapes (in units of k_BT) obtained from the (a) *trans*, (b) *cis* and (c) ligand-free[186] equilibrium simulations of PDZ2, plotted as a function of two essential inter-residue distances. The unlabelled state-like feature at the bottom right of (b) represents weakly populated ($\lesssim 1\%$) sub-regions of states **2** and **5**. (d) Histogram of the state populations in *trans* and *cis* equilibrium, revealing the ligand-induced population shift of PDZ2. (e) Comparison of minimum-energy structures the of states **1**, **2** and **5**, revealing an increased opening of the ligand binding pocket by a downward motion of α_2. (f) Structures of states together with position densities of the ligand. The isosurface encloses a volume with a minimal probability of 0.4 to find a ligand atom within in all simulation snapshots belonging to a specific state. Fixed points for the comparison are the C_α atoms of strands β_4 and β_6.

inherent property of the protein's rugged free energy landscape,[116,117] or are these features rather induced by the ligand? Figure 5.4c addresses this question by showing the free energy landscape obtained from previously performed $6\times1\mu$s-long simulations of PDZ2 *without* a ligand.[186] While the state separation along coordinate $d_{4,55}$ still exists, we find that states **1**, **2** and **5** merge into a single energy minimum. It is centered at the position of state **2**, but is wide enough to cover a large part of states **1** and **5**. Similarly states **3** and **4** form a weakly populated (2 %) single minimum. This indicates that ligand-free PDZ2 provides the flexibility to assess the entire free energy landscape explored during binding and unbinding, while the interaction with the ligand appears to stabilize conformational states **1** and **4**.

Showing protein structures of the main states together with position densities of the ligand, Fig. 5.4f illustrate these interactions (see also Supplementary Fig. S10). For one, we notice that the opening and closing of the binding pocket (described by $d_{20,71}$) is associated with the conventional binding of the ligand's C-terminus in this pocket, which stabilizes closed state **1** in *trans*. In the open state **2**, the probability to find the ligand in

its binding mode is significantly decreased, pointing to a reduced ligand affinity of the protein. On the other hand, we find that the distinct conformations of the protein's termini described by $d_{4,55}$ are a consequence of the formation of contacts with the ligand's N-terminus in states **3** and **4**, which are absent in states **1**, **2** and **5**. In particular, state **5** represents a situation where the hydrophobic photoswitch of the ligand forms a contact with a hydrophobic bulge at the protein surface around Ile20, which can be classified as unspecific binding of the ligand to the protein surface.

Adopting our *trans*-to-*cis* non-equilibrium simulations, we can describe the overall structural evolution of PDZ2 in terms of time-dependent expectation values of various observables. As an example, Figs. 5.5a,b show the time evolution of the two C_α-distances $d_{20,71}$ and $d_{4,55}$ introduced above. Following *trans*-to-*cis* ligand switching, it takes about 100 ns until the sub-picosecond photoisomerization of the photoswitch affects the protein's binding region (indicated by $d_{20,71}$), which becomes wider as the ligand moves out. The flexible N-terminal region indicated by $d_{4,55}$, on the other hand, undergoes conformational changes already within a few nanoseconds. The weak correlation between the two inter-residue distances (i.e., $\langle d_{20,71} d_{4,55}\rangle (\langle d_{20,71}^2\rangle \langle d_{4,55}^2\rangle)^{-1/2} \lesssim 0.02$ for all data), however, indicates that this early motion of the terminal region may be not directly related to the functional dynamics of PDZ2. Interestingly, the associated root mean squared deviations (RMSD) of the two distances show quite similar behavior. Moreover, Supplementary Fig. S11 displays various ligand-protein distances and contact changes, which illustrate that the ligand leaves the binding pocket on time-scales of $0.1 - 1\,\mu$s. When we calculate the dynamical content of all considered intraprotein C_α-distances, we obtain a time-scale distribution that roughly resembles the experimental result (Fig. 5.3g,h).

It is instructive to consider the resulting time-dependent populations of the protein's metastable states. Choosing initial conditions close to the crystal structure,[188] Fig. 5.5c exhibits the *trans*-to-*cis* time evolution of the state populations. The system starts at time $t = 0$ almost completely in state **1** and converts to the other states within microseconds. To rationalize these findings, we construct a Markov state model[68,306] (MSM) which describes the conformational dynamics of PDZ2 via memory-less jumps between metastable states. To this end, we calculate a transition matrix T containing the probabilities T_{ij}, that the system jumps from state i to j within lag time τ_{lag}, and determine its eigenvectors ψ_k and eigenvalues λ_k (see Supplementary Methods and Fig. S12 for technical details). As a first impression, Figs. 5.5c,d compares the state populations obtained from the non-equilibrium MD simulations and the corresponding MSM predictions (using $\tau_{\text{lag}} = 1\,$ns). We find excellent agreement for the first three decades of time, but only qualitative agreement in the last decade, which reflects the bias of our non-equilibrium MD simulations towards shorter time-scales ($75 \times 1\,\mu$s-long and $25 \times 10\,\mu$s-long data). Showing a network representation of the MSM, Fig. 5.5e illustrates the connectivity and transition times of the system. We see that the open-close transition of the binding

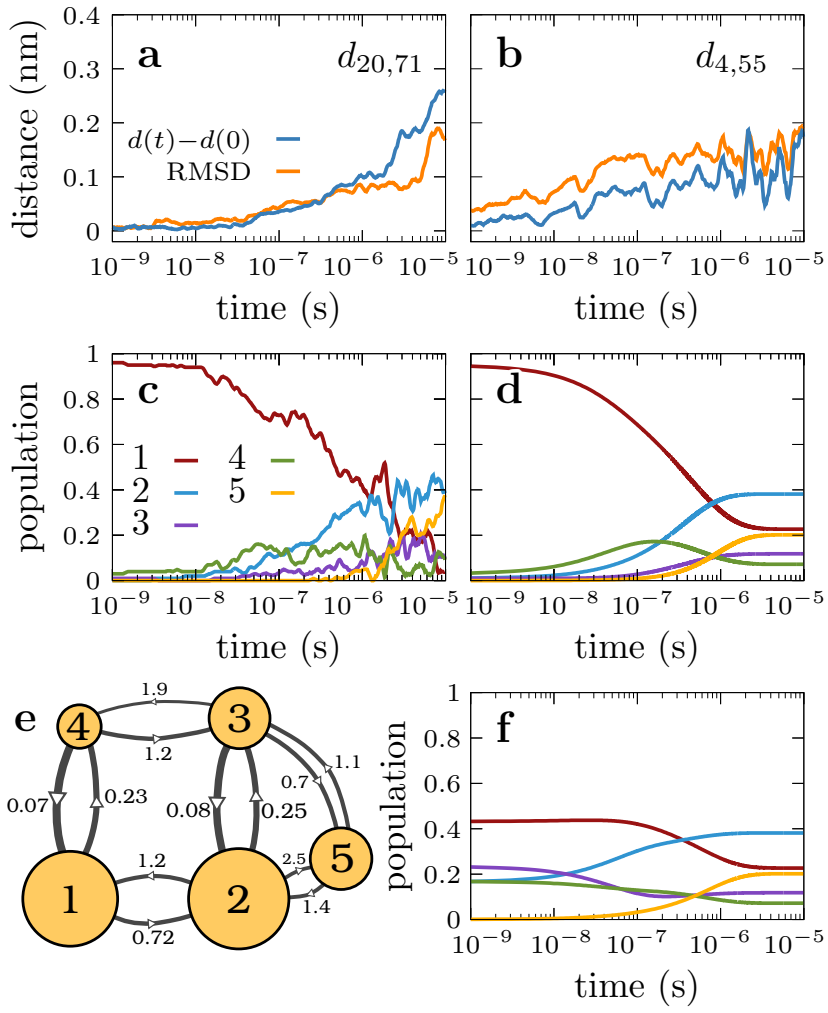


FIGURE 5.5: Time evolution of various structural descriptors, following *trans*-to-*cis* ligand-switching of PDZ2. Shown are means (blue) and RMSD (orange) of C_α-distances (a) $d_{20,71}$ and (b) $d_{4,55}$, as well as (c,d,f) populations of conformational states. For easier representation, all MD data were smoothed. Starting at time $t = 0$ almost completely in state **1**, we compare results from (c) the non-equilibrium MD simulations to (d) the corresponding predictions of a Markov state model (MSM). (e) Network representation of the MSM. The size of the states indicate their population, the thickness of the arrows and numbers indicate the transition times (in μs). For clarity, we discard transitions that take longer than 2.5 μs. (f) MSM simulations of the *trans*-to-*cis* transition, using *trans* equilibrium initial conditions.

pocket occurs on a time-scale of $\sim$ 1 μs, whereas transitions from states **1** and **2** to states **4** and **3** are a factor 4 faster with a back-rate that is even a factor 10 faster.

Assuming a time-scale separation between fast intrastate fluctuations and rarely occurring interstate transitions, MSM theory[68] states that the time-dependent expectation value of any dynamical observable can be written as a sum over exponential functions e^{-t/t_k} weighted by the projection of the observable onto the kth eigenvector of transition matrix T. The implied time-scales $t_k = -\tau_{\text{lag}} / \ln \lambda_k$ of the MSM therefore govern the time

evolution of such different observables as vibrational spectra and state populations.[56] To facilitate a comparison of experimental and simulated time evolutions, we run a MSM simulation using *trans* equilibrium initial conditions, which is also the starting point of the *trans*-to-*cis* experiments. Comparing the simulation results (Fig. 5.5f) to the experimental time traces (Fig. 5.3), we find that both spectral and population evolutions appear to be completed on microsecond time-scale. Moreover, the MSM populations exhibits various transient features on time-scales of 10 – 100 ns, which are also present in the experimental time signals.

5.3 Discussion and Conclusions

Combining transient IR spectroscopy and non-equilibrium MD simulations, we have described the ligand-induced conformational transition in the PDZ2 domain, which is thought to be responsible for protein allosteric communication. We have found that the free energy landscape of PDZ2 can be described in terms of a few metastable states with well-defined structure (Fig. 5.4), although the mean structural changes upon ligand switching are rather small. That is, the secondary and tertiary structure of the protein are quite similar ($\lesssim$ 1 Å RMS displacement) in the different states, and only modest ($\sim$ 20 %) shifts of the state's population are found (Fig. 5.4b). On average, the measurable structural change is therefore only in the order of 0.2 Å. In light of this result, it is remarkable that we can observe such minor structural changes by transient IR spectroscopy (Fig. 5.2), unpinning the extraordinary structural sensitivity of the method.

Using isotope labeling to discriminate the dynamics of protein and ligand, the resulting time-resolved double-difference IR spectra have revealed complex kinetics of the protein that cover many time-scales (Fig. 5.2). The spectra for *trans*-to-*cis* and *cis*-to-*trans* ligand-switching are not mirror-images from each other, and the *trans*-to-*cis* signals exhibit short-time transients that are not found for *cis*-to-*trans*. Moreover, the *cis*-to-*trans* transition does not seem to be finished within 42 μs (Fig. 5.2c). The overall slower response of the *cis*-to-*trans* transition reflects the general observation that enforced leaving of a well-defined (low entropy) ligand binding structure (here *trans*) occurs faster than starting in a conformationally disordered (high-entropy) state (here *cis*) and trying to find stabilizing interactions to end in a more organized structure.[307]

More specifically, the *trans*-to-*cis* non-equilibrium simulations reveal that the ligand remains bound with its C-terminus to the protein binding site between β_2 and α_2 up to about 1 μs. In this way, it stabilizes the main bound protein conformation (state **1**). At longer times, it starts to move out from the binding pocket, but remains non-specifically bound to the protein surface. While diffusion on the surface may continue for long times after *trans*-to-*cis* switching, it only little affects the protein internal structure. Nevertheless, this diffusion will be the first rate-limiting step after *cis*-to-*trans* switching, which

might be the reason that the ligand does not completely localize in the binding pocket within 42 μs.

The existence of well-defined metastable conformational states implies a time-scale separation between fast intrastate fluctuations and rarely occurring interstate transitions. This allowed us to construct a Markov state model (MSM), which illustrates the connectivity and transition times between the metastable states (Fig. 5.5d). In particular, the discrete time-scales predicted by the MSM are directly reflected in the dynamical content calculated for experiments and MD simulations (Fig. 5.3e-h), which both cover time-scales from $\sim$1 ns to 10 μs. Reflecting different observables (transition dipole vs. C_α-distances, respectively), the weights of the various peaks are different.

While ligand switching was shown to cause a conformational transition of PDZ2 in terms of the mean structure, at the same time it may also effect a change of the protein's fluctuations. Comparing the time evolution of the means of the distances and their RMSD, Figs. 5.5a,b reveal that the two quantities correlate closely, a behavior that is found for all considered C_α-distances (Supplementary Fig. S7). This finding reflects the fact that the C_α-distance distributions pertaining to the individual states are in most cases well separated (Supplementary Fig. S13), such that a transition between two states affects both mean and variance. Accounting for an entropic contribution of the conformational transition, the latter effect is often referred to as "dynamic allostery".[26,112,113] The above findings indicate that allosteric transitions may involve both, conformational and dynamic changes in the case of the PDZ2 domain.[278] The answer to what is the dominant effect will greatly depend on the system under consideration and on the applied experimental method. While the overall structural change ($\lesssim$ 1 Å RMS displacement) may be too small to be detected by structure analysis, NMR relaxation methods can sensitively explore the structural flexibility of proteins. The IR spectrum of the amide I band, in contrast, is commonly thought of as a measure of structure,[298] but dephasing due to fast fluctuation might also affect the IR lineshape.

In conclusion, we have characterized the non-equilibrium allosteric transition of a widely studied model system, the PDZ2 domain, using a joint experimental-theoretical approach. The protein domain *per se* was kept unmodified, hence we believe that the concept of ligand-switching mimics very closely the naturally occurring allosteric perturbation caused by ligand (un)binding events. While the ligand-induced allosteric transition originates from a population shift between various metastable conformational states, the measurable mean structural change of the protein may be tiny and therefore difficult to observe.[278] Moreover, we suggest that the separation between purely dynamically driven allostery and allostery upon a conformational change may not be as clear-cut as previously thought, but rather that there may be an interplay between both that allows proteins to adapt their free energy landscape to incoming signals.

Supplementary Information

5.4 Materials and Methods

5.4.1 Protein and Peptide Preparation

Expression of the wild type PDZ2 domain from human phosphatase 1E,[114] isotope labelled ($^{13}C^{15}N$) protein variant and synthesis of the photoswitchable peptide ligand was performed as described earlier.[16,159] The wild type RA-GEF-2 sequence was modified in order to enable cross-linking the photoswitch, while preserving residues that are important for regulation and binding. That is, amino acids at positions (-1) and (-6) were chosen as anchoring points for the photoswitch and mutated into cysteine residues. Four N-terminal residues (RWAK) were added to the sequence in order to improve the water solubility and facilitate the concentration determination of the construct. Final sequence of the peptide was RWAKSEAKECEQVSCV.

The purity of all samples was confirmed by mass spectrometry analysis (Fig. S1). All samples were dialyzed against 50 mM borate, 150 mM NaCl buffer, pH = 8.5. For transient infrared measurements, samples were lyophilized and resuspended in D_2O. Incubation of the samples in D_2O overnight at room temperature before the measurements eliminated H/D exchange during experiments. The concentration of the samples was determined via the tyrosine absorption at 280 nm for the protein and 310 nm for the peptide and confirmed by amino-acid analysis.

5.4.2 Determining the Binding Affinity

Isothermal titration calorimetry (ITC) measurements were performed on a MicroCal ITC200 (Malvern, UK). In order to ensure the obtained values for the *cis* and *trans* measurement were mutually comparable, the experiments were performed using the same stock solution of the peptide and protein for both measurements, and under exactly the same experimental conditions. The experiment was performed in triplicate in order

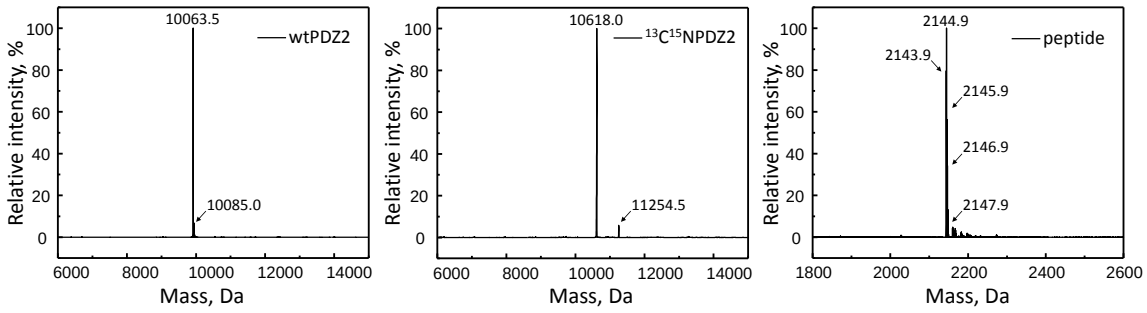


FIGURE S1: Mass spectra for the wild type PDZ2 domain (left), isotopically labelled PDZ2 domain (center) and photoswitchable peptide (right).

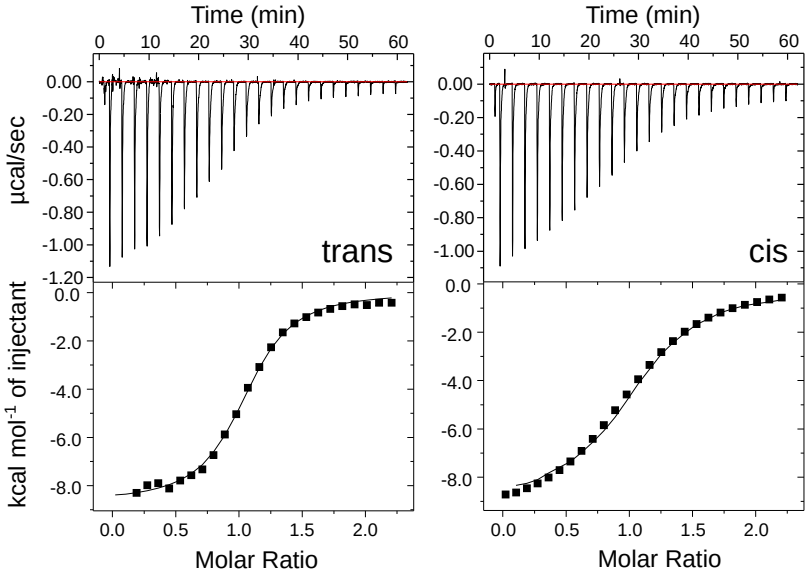


FIGURE S2: ITC thermograms of the wild type PDZ2 domain with the photoswitchable peptide in the *trans* and *cis* configurations (left and right plot, respectively).

to ensure the reproducibility of the data. The sample cell was loaded with 250 μl of 80 μM PDZ2 domain solution and the syringe was loaded with 40 μl of 800 μM photoswitchable peptide solution. For the *trans* measurement, the system was kept in the dark for the duration of the experiment, while for the *cis* measurement the syringe was constantly illuminated with a 370 nm cw laser (CrystaLaser, power $\approx$ 90 mW).[295] The results are shown in Fig. S2.

As alternative method to determine the binding affinity, we also used circular dichroism (CD) spectroscopy as well as fluorescence quenching. Both spectroscopic signals change upon the formation of a protein-ligand complex, hence, when measuring them in dependence of peptide and protein concentration, the binding affinity can be fitted assuming a bimolecular equilibrium. CD measurements were done on Jasco (Easton, MD) model J810 spectropolarimeter in a 0.1 cm quartz cuvette as described previously.[295] Intrinsic tryptophan fluorescence quenching experiment was done on PelkinElmer spectrofluorimeter as described previously.[295] In either case, the protein concentration was kept constant at 5 μM, respectively, while the peptide concentrations were varied. Fig. S3 shows the results for the CD spectroscopy and trypthophan fluorescence quenching, while Table S1 compares the binding affinities obtained from all different methods.

TABLE S1: Comparison of K_d values obtained from ITC, CD and intrinsic fluorescence quenching.

Kd (μM)	*trans*	*cis*
ITC	2.7 ± 0.2	9.1 ± 0.4
CD	1.6 ± 0.1	10.1 ± 0.3
Fluorescence	1.7 ± 0.3	9.8 ± 1.2
Mean	2 ± 0.6	9.7 ± 0.5

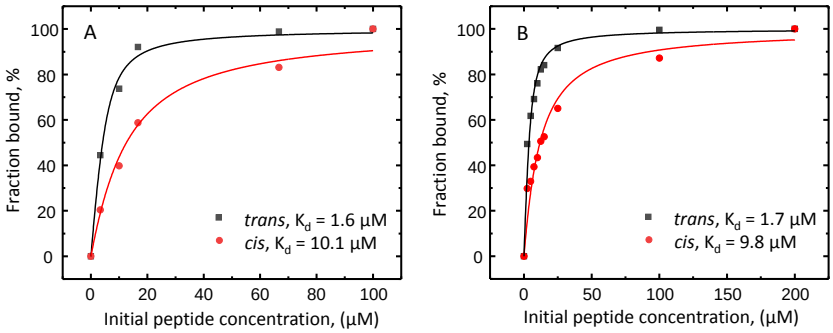


FIGURE S3: Binding affinity determination for the wild type PDZ2 domain and photoswitchable peptide in cis (red) and trans (black) configuration of the peptide determined by CD (left) and intrinsic tryptophan fluorescence quenching (right).

5.4.3 Transient IR Spectroscopy

Transient VIS-pump-IR-probe spectra were recorded using two electronically synchronized Ti:Sapphire laser systems[296] running at 2.5 kHz. The wavelength of the pump-laser was tuned as to obtain 380 nm pump pulses (2.1 μJ) for the *trans*-to-*cis* experiment, and 420 nm (1.3 μJ) for the *cis*-to-*trans* experiment, respectively, *via* second harmonic generation in a BBO crystal. The beam diameter of the pump pulse at the sample position was ≈180 μm, employing a pulse duration of ≈200 ps (by extracting the light directly after the regenerative amplifier and before the compressor) to minimize the sample degradation during the measurements. Mid-IR probe pulses centered at ≈1630 cm^{-1} (pulse duration ≈100 fs, beam diameter on the sample ≈150 μm) were obtained in a optical parametric amplifier (OPA),[198] passed through a spectrograph and detected in a 2×64 MCT array detector with a spectral resolution of ≈2 cm^{-1}/pixel. Pump-probe spectra were acquired up to the maximum delay value of ≈42 μs with a time resolution of ≈200 ps. Normalisation for noise suppression was performed as described in Ref..[297]

The samples (≈700 μl) were pumped through a closed flow-cell system purged with N_2. The system consisted of a sample cell with two CaF_2 windows separated by a 50 μm Teflon spacer and a reservoir. The flow speed in the sample cell was optimized in order to minimize loss of sample at the largest pump-probe delay time (≈42 μs) on the one hand, but to have the sample exchanged essentially completely for the subsequent laser shot after 400 μs on the other hand. The concentrations of the samples were set at 1.25 mM for the peptide and 1.5 mM for the protein. A slight excess of protein was needed to ensure that the peptide was fully saturated with the protein; in order to eliminate the response of free, photoswitchable peptide. As a reference, FTIR difference spectra have been taken in a Bruker Tensor 27 FTIR spectrometer, using the same sample conditions.

For the experiment with *trans*-to-*cis* switching, we relied on thermal *cis*-to-*trans* back

reaction. By comparing its rate with the isomerization probability induced by the 380 nm pump light (determined by pump light power, total sample volume, absorption cross sections,[115] and isomerization quantum yield[299]), we estimated that the photo-equilibrium in the total sample volume is 70%/30% *trans*/*cis* during measurement. It furthermore helps that the absorption cross section at 380 nm of the azobenzene moiety in the *trans*-state is ≈20 times larger than that of the *cis*-state,[115] which leads us to conclude that >97% of the molecules in the *trans*-to-*cis* experiment undergo the desired isomerisation direction. For the experiment with *cis*-to-*trans* switching, the sample could be actively switched back by illuminating the reservoir with an excess of light at 370 nm from a cw laser (CrystaLaser, 150 mW).

5.4.4 Time-Scale Analysis

To determine "time-scale spectra", the transient infrared data were analyzed by a maximum entropy inversion of the Laplace transform.[301–303] That is, the data were essentially fit to a multiexponential function:

$$S(\omega_i, t) = a_0(\omega_i) - \sum_i a(\omega_i, \tau_j) e^{-t/\tau_j} \tag{S2}$$

where the index j refers to a kinetic component with time constant τ_j, and i to a probe frequency ω_i. The time constants τ_j were fixed and distributed equidistantly on a logarithmic scale with 10 terms per decade, while the amplitudes $a(\omega_i, \tau_j)$ were the free fitting parameter. However, inversion of Eq. S2 directly is an ill-posed problem, hence a penalty function was added to the RMSD of the fit, maximizing the Shannon entropy of the amplitudes $a(\omega_i, \tau_j)$ at the same time. A special feature of the algorithm used here accounts for the fact that the amplitudes $a(\omega_i, \tau_j)$ may be positive or negative, see Ref.[303] for details. Fig. S4 shows the complete time-scale spectrum in a 2D representation, while the blue lines in Fig. 3(a-d) show cuts for certain probe frequencies ω_i. The "dynamical content" $D(\tau_i)$ was then calculated as:

$$D(\tau_j) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} a^2(\omega_i, \tau_j)} \tag{S3}$$

where the sum runs over all frequency positions of the transient infrared signal.

In recent works,[14] we had applied a similar approach, utilizing as a penalty function $\sum_j \left(a(\omega_i, \tau_{j+1}) - a(\omega_i, \tau_j)\right)^2$ for a smooth time-scale spectrum[308] instead of the Shannon entropy. We found that both approaches reveal mutually consistent results, but that the resolution in the time-scale spectra is higher with the maximum entropy method.

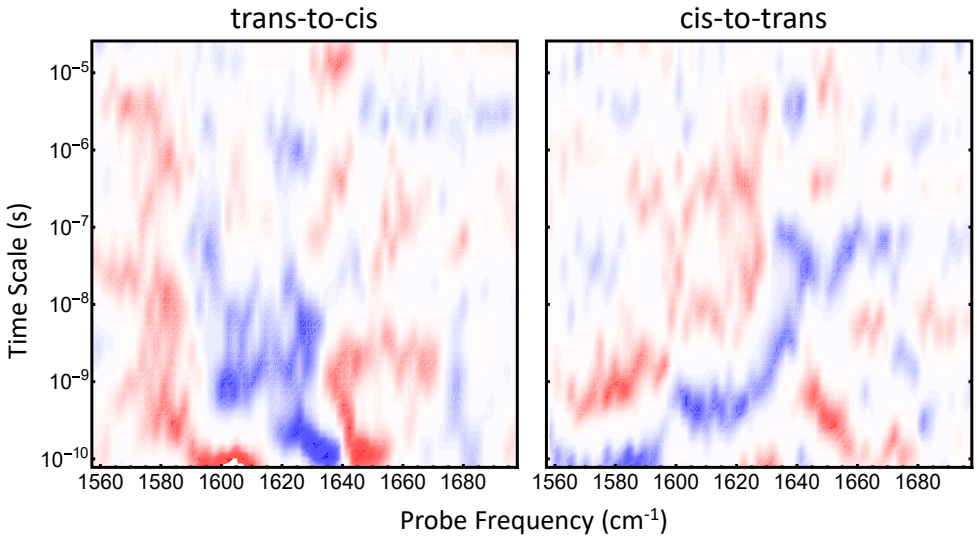


FIGURE S4: Time-scale analysis $a(\omega_i, \tau_j)$ of the $^{13}C^{15}N$-WT difference data for *trans*-to-*cis* (left) and *cis*-to-*trans* (right) switching. Red colours indicate positive amplitudes $a(\omega_i, \tau_j)$, blue colors negative amplitudes $a(\omega_i, \tau_j)$.

5.5 Computational methods

5.5.1 MD simulations

All MD simulations of PDZ2 were performed using the GROMACS v2016 software package[48] and the Amber99*ILDN force field.[139–141] Force field parameters of the azobenzene photoswitch were taken from Ref..[159] Protein-ligand structures were solvated with ca. 8000 TIP3P water molecules[148] in a dodecahedron box with a minimal image distance of 7 nm. 16 Na^+ and 16 Cl^- were added to yield a charge-neutral system with a salt concentration of 0.1 M. All bonds involving hydrogen atoms were constrained using the LINCS algorithm,[206] allowing for a time step of 2 fs. Long-range electrostatic interactions were computed by the Particle Mesh Ewald method,[309] whereas the short-range electrostatic interactions were treated explicitly with the Verlet cutoff scheme. The minimum cutoff distance for electrostatic and van der Waals interactions was set to 1.4 nm. A temperature of 300 K was maintained via the Bussi thermostat[146] (aka velocity-rescale algorithm) with a coupling time constant of τ_T = 0.1 ps. A pressure P =1 bar was controlled using the pressure coupling method of Berendsen[147] with a coupling time constant of τ_P = 0.1 ps.

The starting structure of the photoswitched ligand bound to PDZ2 was prepared previously (see Ref.[159]) based on the crystal structure (PDB ID 3LNX[188]). Here, the azobenzene photoswitch was attached in *trans* conformation to the ligand at positions (-6) and (-1), which had been mutated to cysteins as in experiment to provide covalent connection points. Residues missing at the N-terminus of the ligand were added (see Sec. 5.4.1). Following NPT equilibration of the system in *trans* conformation for 10 ns, 4 statistically independent (i.e., with different initial velocity distributions) NVT runs

of 100 ns each were performed. For one, we selected 5 randomly chosen snapshots from the end of these trajectories to perform $5 \times 5\mu$s-long ***trans* equilibrium** simulations. Moreover, we selected 25 randomly chosen snapshots from each of the last 50 ns of these four NVT trajectories to perform *trans*-to-*cis* nonequilibrium simulations, yielding a total of 100 starting structures which consists mostly of metastable state 1 (for state definition, see Sec. 5.5.2). Employing these initial conditions, *trans*-to-*cis* photoswitching was performed using a previously developed potential-energy surface switching approach.[40] All 100 ***trans*-to-*cis* nonequilibrium** simulations were run for 1μs; 25 of them were extended to a length of 10 μs.

Upon switching the ligand from *trans* to *cis* configuration, PDZ2 undergoes a nonequilibrium time evolution until it relaxes within a few microseconds (see below) into its *cis* equilibrium state, describing the unbound protein-ligand complex. Performing $25 \times 10\,\mu$s-long *trans*-to-*cis* nonequilibrium simulations, we took the last 7μs of each trajectory to estimate the rather heterogeneous conformational distribution of the ***cis* equilibrium** state. To generate initial structures for *cis*-to-*trans* photoswitching, we took from the 25 *trans*-to-*cis* trajectories 100 randomly chosen snapshot at a simulation time around 3.0μs. Following photoswitching, 100 ***cis*-to-*trans* nonequilibrium** trajectories were simulated for a trajectory length of 1 μs; 10 simulations were extended to a length of 8 μs.

Gromacs tools *gmx angle* and *gmx mindist* were employed to compute backbone dihedral angles, interresidue C_α-distances, and the number of contacts between various segments of PDZ2. Time-dependent distributions and mean values of these observables were calculated via an ensemble average over 100 nonequilibrium trajectories.

5.5.2 Dimensionality reduction and clustering

Choice of input coordinates. We wish to obtain a common set of collective variables that describe the structural dynamics of all simulations of PDZ2, including *cis* and *trans* equilibrium simulations as well as *trans*-to-*cis* and *cis*-to-*trans* simulations (totaling ∼510 μs simulation time). In a first step, we need to choose suitable internal coordinates that account for the conformational transitions of the system.[53] To this end, we determined 56 C_α-distances $d_{i,j}$ between residues i and j that are not redundant (such as $d_{i,j}$ and $d_{i,j\pm1}$) and whose ensemble average changes significantly ($\langle d_{ij} \rangle \geq 0.5\,$Å) during the first microsecond *trans*-to-*cis* nonequilibrium simulations, see Fig. S5. Moreover, we considered all backbone dihedral angles that show a change of $\gtrsim 10^\circ$ from their initial value during the *trans*-to-*cis* nonequilibrium simulations.

Principal component analysis. Since the interresidue C_α-distances appear to provide more information, these coordinate are chosen for the subsequent principal component analysis (PCA), which was performed on all data.[53] For adequate relative weighting of

short and long distances, the data was normalized.[52] Diagonalizing the resulting covariance matrix, we obtain its eigenvectors (yielding the PCs) and eigenvalues (reflecting the fluctuations of the PCs). The first two PCs cover 43 % of the overall fluctuations, while six PCs yield about 65 %. Calculating the free energy profiles pertaining to the PCs, we find that in particular PC 1–4, 6 and 7 show multistate behavior reflecting metastable states.

Density-based clustering. Including these 6 PCs, we performed robust density-based clustering,[54] which first computes a local free energy estimate for every structure in the trajectory by counting all other structures inside a 6-dimensional hypersphere of fixed radius R. Normalization of these population counts yields densities or sampling probabilities P, which give the free energy estimate $\Delta G = -k_{\mathrm{B}}T \ln P$. Thus, the more structures are close to the given one, the lower the free energy estimate. By reordering all structures from low to high free energy, finally the minima of the free energy landscape can be identified. By iteratively increasing a threshold energy, all structures with a free energy below that threshold that are closer than a certain lumping radius will be assigned to the same cluster, until all clusters meet at their energy barriers. In this way, all data points are assigned to a cluster as one branch of the iteratively created tree. For PDZ2, we used a hypersphere $R = 0.579$ that equaled the lumping radius employed in the last step.

Figure S6(top) shows the resulting total number of states obtained as a function of the minimal populations P_{min} a state must contain. Here we chose $P_{min} = 50\,000$, resulting in a clustering into 12 states. According to visual inspection of the resulting free energy landscapes (Fig. S6(middle)), these states separate accurately all density maxima of the system. Since the 5 lowest populated states cover less than 5 % of the total population, we lumped them to main states **1** to **7** as follows: (**1, 9**)→**1**, (**2, 10**)→**2**, (**4, 12**)→**4**, (**5, 8, 11**)→**5**. This is justified due to their geometric vicinity in the free energy landscape (Fig. S6(middle)), as well as due to their kinetic vicinity in the transition matrix. Following the calculation of the time-dependent states populations, in a last step we lumped states (**4, 7**)→**4** and states (**5, 6**)→**5** for the sake of easy interpretability of the figures in the main text.

Identification of essential internal coordinates. Finally we employed a recently proposed machine learning approach[305] to identify the internal coordinates that allow to discuss the 5 main states of PDZ2 in a two-dimensional free energy landscape. On the basis of the decision-tree based program XGBoost,[310] we trained a model that determines the features of the molecular coordinates that are most important to discriminate given metastable states. Using a new algorithm that exploits this feature importance via an iterative exclusion principle, we identified the essential internal coordinates, that is, the most important C_α-distances of PDZ2. Figure S6(bottom) shows that three distances, $d_{20,71}$, $d_{4,55}$ and $d_{27,69}$ suffice to qualitatively distinguish the 5 main states of PDZ2. The

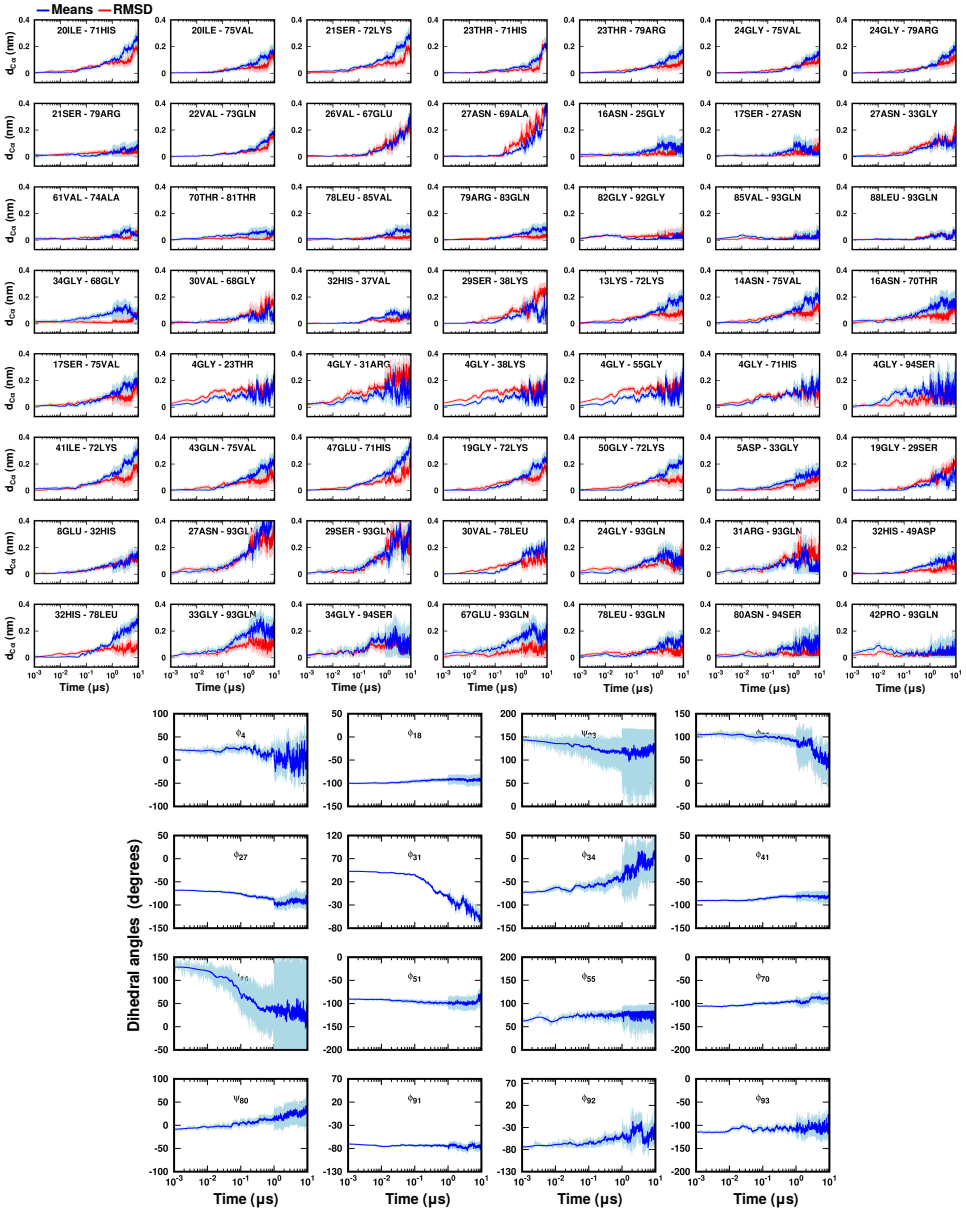


FIGURE S5: Choice of input coordinates. (Top) Time evolution of the average $|d(t) - d(0)|$ and their root mean squared deviation (RMSD) of 56 selected interresidue C_α-distances of PDZ2 upon *trans*-to-*cis* ligand switching. Since the secondary structural elements of PDZ2 are given as β_1 (residues 6 – 12), β_2 (20 – 24), β_3 (35 – 40), α_1 (45 – 49), β_4 (57 – 61), β_5 (64 – 65), α_2 (71 – 80), and β_6 (84 – 90), the first 11 distances report on the binding pocket, distances 12 – 26 report on dynamics localized inside some secondary structure element, and the remaining distances account for distances between two secondary structure element. (Bottom) Time evolution of the average value of all backbone dihedral angles that show a change of $\gtrsim 10^\circ$ from their initial value during the first microsecond. The shaded areas reflect the corresponding RMSD. To avoid problems associated with the circular statistics of the angles, we performed maximal gap shifting[54] of the angles prior to averaging.

XGBoost parameters are chosen as in Ref.,[305] including learning rate $\eta = 0.3$, maximum tree depth of 6, 10 training rounds, and 70% and 30% of the data used for training and

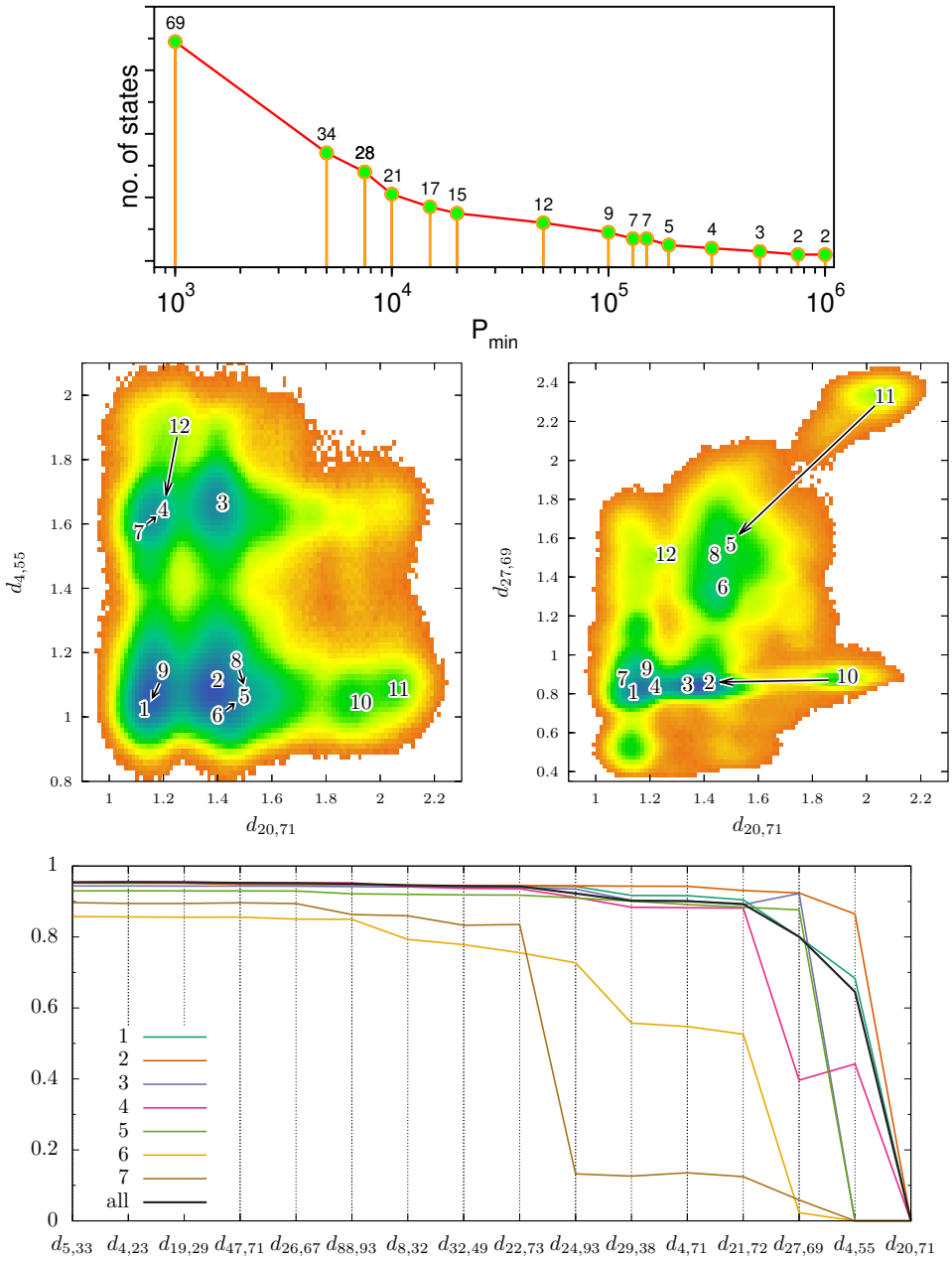


FIGURE S6: Identification of metastable conformational states and essential internal coordinates of PDZ2. (Top) Robust density-based clustering using 6 PCs and all simulation data ($2.5 \cdot 10^7$ points). Shown is the total number of states obtained as a function of the minimal populations P_{min} a state must contain. (Middle) Energy landscape along along C_α-distances $d_{20,71}$, $d_{4,55}$ and $d_{27,69}$, where numbers indicate the center of all 12 metastable states. Arrows indicate the lumping of the 12 microstates into 5 main states. (Bottom) Essential internal coordinates of PDZ2 identified by XGBoost.[310] Shown is the resulting accuracy of the respective model after discarding the distances on the abscissa, in black the total 'multiclass error' N_{correct}/N, in color the accuracy of predicting state n, $N_{n,\text{correct}}/N_n - N_{n,\text{incorrect}}/N$.

validation, respectively.

Characterization of metastable states. Accompanying the analysis in the main text, Fig. S7 compare the contact maps of main states **1** and **2**.

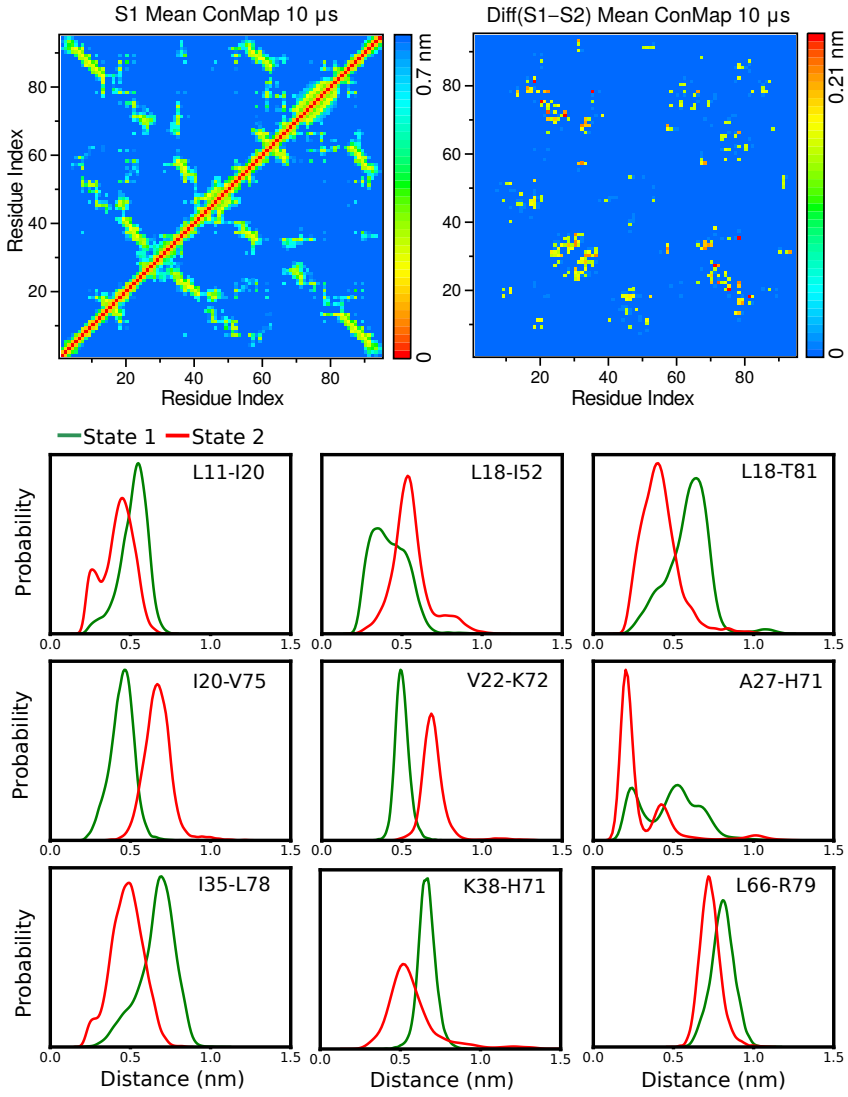


FIGURE S7: (Top) Contact map of metastable state **1**, as well as differences between contact maps of state **1** and state **2**. We consider a contact to be formed when the average minimum distance between two residues is less than 0.45 nm. (Bottom) Distributions of interresidue distances associated with contacts that show a large differences ($\gtrsim$ 0.05 nm) between state **1** and **2**.

5.5.3 Nonequilibrium MD results

Using the initial preparation described in Sec. 5.5.1, we ran *trans*-to-*cis* and *cis*-to-*trans* nonequilibrium simulations of PDZ2. We first consider the *trans*-to-*cis*, where the ligand is initially bound to PDZ2. As representative examples of intraprotein observables, Fig. S5 shows the resulting time evolution of the expectation value of selected C_α-distances (top) and of selected backbone dihedral angels (bottom). Since up to 1µs we average over 100 trajectories, the fluctuations of the data are relatively small. At longer times, however, we have only 25 trajectories to average over, which results in considerably larger fluctuations. Furthermore, at time $t = 1\mu s$ a discontinuity of the observables may occur, since averages over different ensemble sizes may yield different means.

To illustrate the *trans*-to-*cis* time evolution of the ligand, Figs. S8 shows representative structures of the ligand-protein complex. It takes about 1µs for the ligand to unbind from the binding pocket. Alternatively, Figs. S9a,b shows the *trans*-to-*cis* time evolution of the expectation value of ligand-protein distances $d_{(-4),24}$ and $d_{(-6),30}$. The former observable directly reports on the ligand's moving out of the binding pocket. Indicating the distance between the $\beta_2\beta_3$-loop and the ligand, the increase of the latter reflects the overall unbinding of the ligand. Alternatively, the unbinding process can be described by the number of ligand contacts with some protein region. For example, Figs. S9c,d reveals a loss of ligand contacts with β_2 and an increase of contacts with α_2.

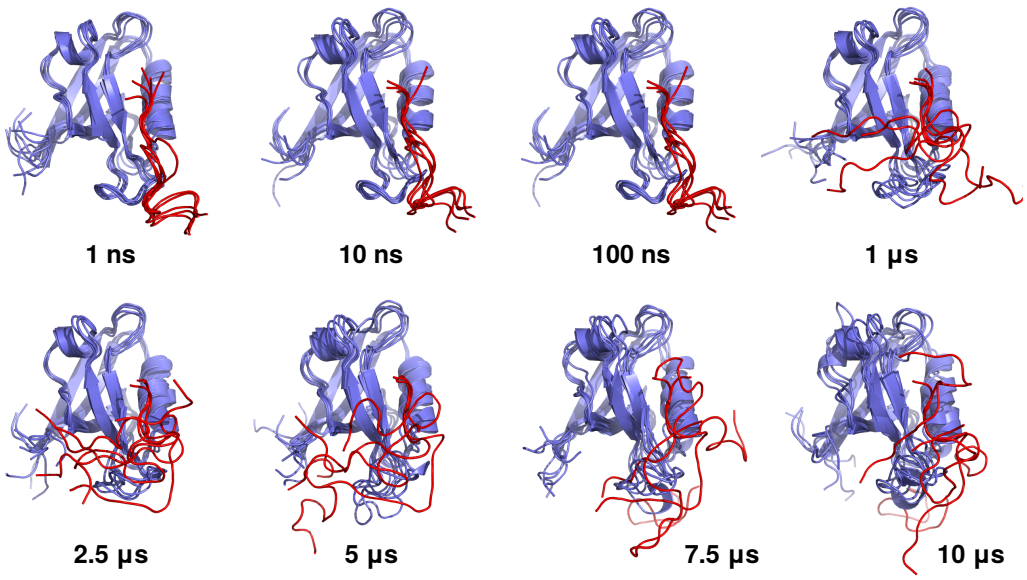


FIGURE S8: Evolution of ligand position in time, obtained from five randomly selected *trans*-to-*cis* nonequilibrium simulations. Protein backbone is shown in blue, ligand backbone in red. The ligand remains bound with its C-terminus to the binding site between β_2 and α_2 up to 1 µs. At longer times, it starts to unbind from the binding pocket, and binds nonspecifically to the protein surface.

So far we have focused the discussion on the dynamics following *trans*-to-*cis* ligand switching. Likewise, we may adopt our *cis*-to-*trans* nonequilibrium simulation to study the reverse process of ligand binding. As representative examples, we consider again the intraprotein C_α-distances $d_{20,71}$ and the ligand-protein distance $d_{(-4),24}$. Unlike to the case of *trans*-to-*cis*, Figs. S9e,f show that for *cis*-to-*trans* both distances remain almost constant (and similar for all other considered distances). This finding reflects the very inhomogeneous *cis* initial state, which is very difficult to sample.

As an estimate of the sampling error, we performed a bootstrapping analysis where, for a given sample of 100 points at time t, 1000 bootstrap samples were considered. The resulting error bars (shown in black) reflects a confidence interval of 68%. As expected, we find that all observables obtained for the *trans*-to-*cis* transition (blue lines) are significantly better sampled than the ones obtained for the *cis*-to-*trans* transition (red lines). For this reason, we use the *cis*-to-*trans* data only to calculate the highly averaged (and thus fairly robust) dynamical content [Eq. (S3)], but otherwise focus in the main text on the relatively well sampled *trans*-to-*cis* case.

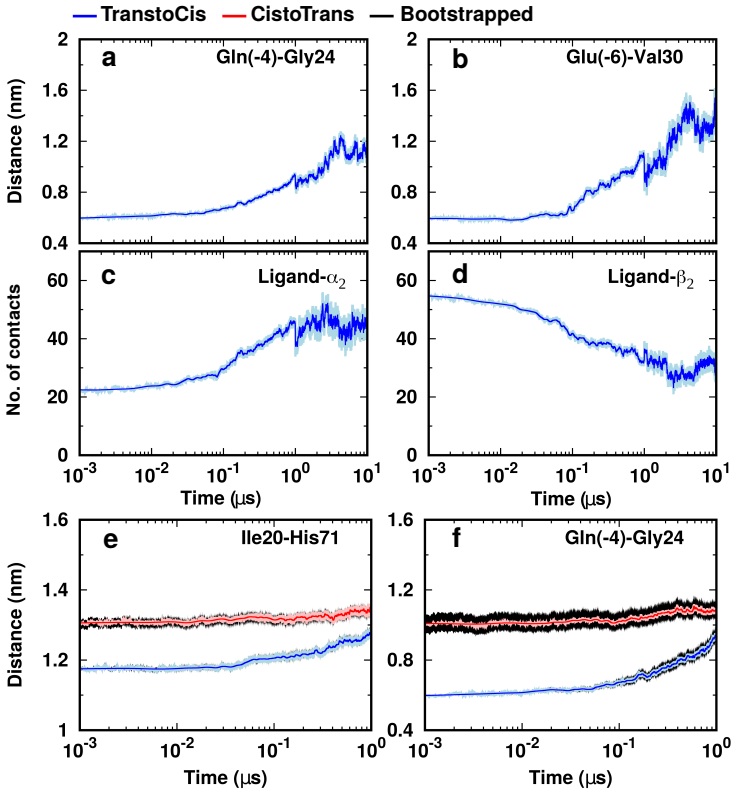


FIGURE S9: *Trans*-to-*cis* time evolution of (top) mean ligand-protein distances $d_{(-4),24}$ and $d_{(-6),30}$, as well as (middle) the number of ligand contacts with α_2 and β_2. (Bottom) Response of PDZ2 upon *trans*-to-*cis* (blue lines) and *cis*-to-*trans* (red lines) ligand switching. Black error bars are estimated via bootstrapping, see text.

5.5.4 Markov state model

On the basis of the above defined 7 metastable states, we constructed a Markov state model[68] of the *trans*-to-*cis* transition of PDZ2, using all ($75\times1\,\mu$s and $25\times10\,\mu$s) *trans*-to-*cis* nonequilibrium trajectories. A general problem with the definition of metastable states is that, due to the inevitable restriction to a low-dimensional space combined with insufficient sampling, we often obtain a misclassification of sampled points in the transition regions, which causes intrastate fluctuations to be mistaken as interstate transitions. As a simple but effective remedy, we use dynamical coring which requires that a transition must a minimum time τ_{cor} in the new state for the transition to be counted.[55,311] A suitable quantity that reflects these spurious crossings is the probability $W_i(t)$ to stay in state i for duration t (without considering back transitions). As shown in Fig. 5.5, without coring we observe a strong initial decay of $W_i(t)$ for all states, instead of a simple exponential decay we would expect for Markovian states. Applying coring with increasing coring times, this initial drop vanishes because fluctuations on timescales $t \lesssim \tau_{\mathrm{cor}}$ are removed. Here we determined $\tau_{\mathrm{cor}} = 1\,\mathrm{ns}$ as shortest coring time, which removes the spurious interstate transitions. Figure S10 shows the resulting implied timescales and eigenvectors of the model. Using a lag time of 1 ns, we moreover show the time evolution of the state populations, assuming that we start completely in a specific state.

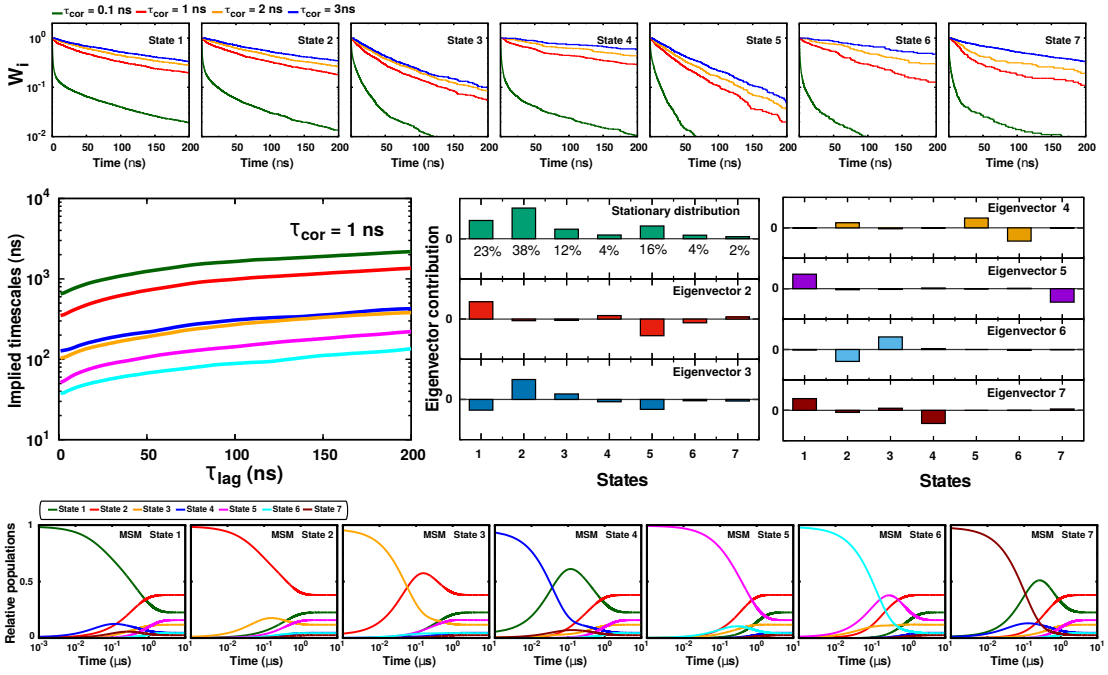


FIGURE S10: Markov state model of the *trans*-to-*cis* transition of PDZ2, employing 7 metastable states. (Top) Probability $W_i(t)$ to stay in state i for duration t, using various coring times τ_{cor}. (Middle) Implied timescales and eigenvectors of a Markov state model using $\tau_{\mathrm{cor}} = 1\,\mathrm{ns}$. (Bottom) Time evolution of the state populations.

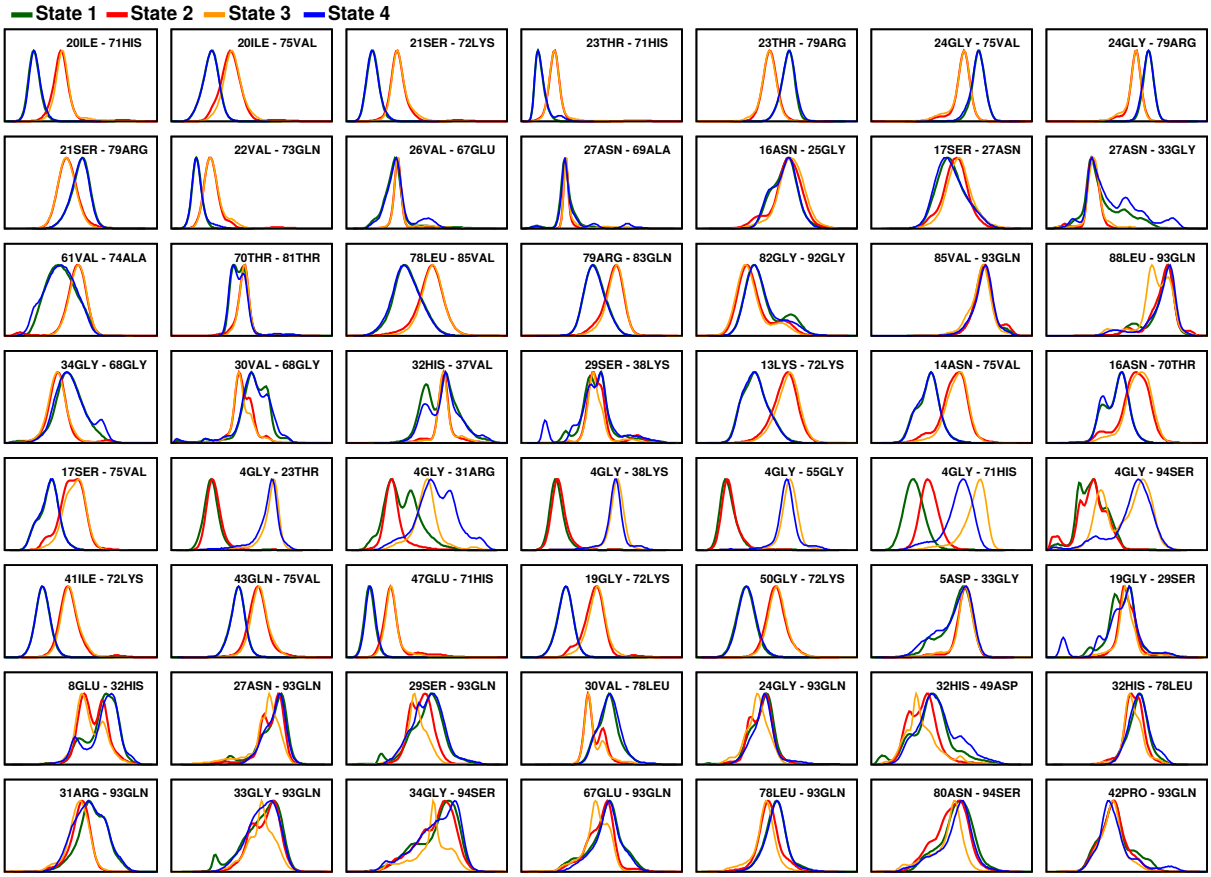


FIGURE S11: Distributions of interresidue C_α-distances, as obtained for the first four metastable states.

5.6 Coupled dynamical and structural allostery in PDZ3 domain

Dynamic interactions regulate the range of biological and biochemical mechanisms. Better mechanistic understanding of these regulation phenomenon may provide the vital knowledge about life at cellular level for example the growth and asymmetry of cellular components.[312] Allostery is known to be among one of the many regulatory mechanisms involved in wide range of biological processes. Transcription regulation, post-translational modification and autoinhibition are wholly or partially dependent on the process of allosterism because all these processes require a stimulus for consequential effects. The changes caused by binding or unbinding of stimuli to the target biomolecules induce conformational changes. These conformational changes are naturally composed up of both dynamical and structural adjustments. In sections 5.1-5.3 we characterized the allosteric response of ligand switched PDZ2L as a remodelling of its rugged free energy landscape, with a few number of structurally well defined states and dynamics distributed over four decades starting from a nanosecond to microseconds timescales. In that case, however, $\approx$ 3.5–folds reduction in binding affinity was recorded upon *trans–to–cis* photoswitching and no major changes in any contact were observed.

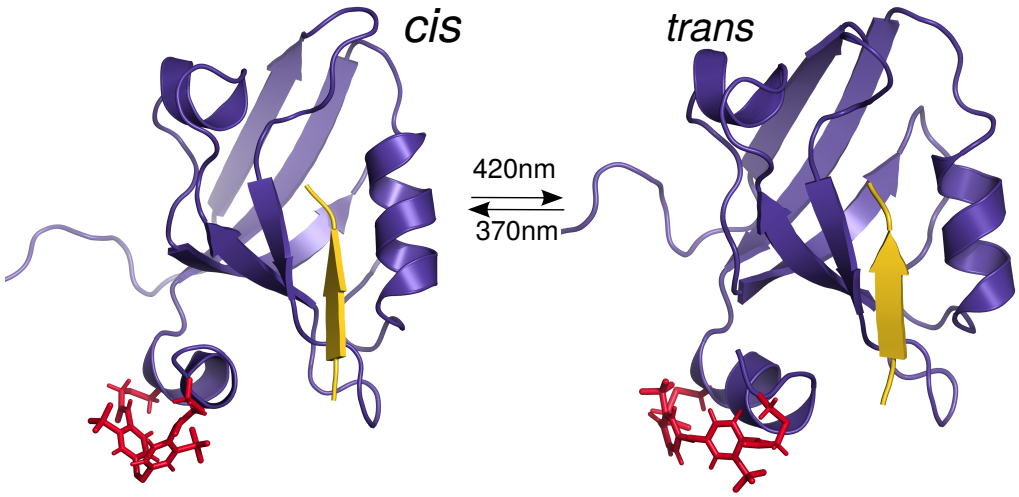


FIGURE 5.6: Cartoon structure of PDZ3 domain in purple; ligand in yellow, photoswitch in red at α_3 and in *cis* and *trans* conformations.

To further investigate the allosteric mechanism we designed an α_3–switched PDZ3 (PDZ3S) in such a way that *cis–to–trans* photoswicthing of α_3 reduces the peptide binding to PDZ3 domain. In addition to the conserved fold of 5-6 β-sheets and 2 α-helices, PDZ3 contains an additional third α-helix (α_3), which packs against the core domain, located

on the opposite side from the binding pocket (see figure 5.6).[121,122] The pioneer experiments of Petit et al., has shown that the truncation of α_3 helix in PDZ3 reduces the binding affinity of a CRIPT ligand by 21–folds.[113] In fact, the experiments of Hamm and coworkers[313] reveal a reduction of $\approx$ 25–folds upon *cis–to–trans* α_3–swicthing of PDZ3S. To provide the structural and dynamical basis of this astonishing reduction in binding affinity we preformed ($99 \times 1\ \mu s$) *cis-to-trans* nonequilibrium and ($4 \times 1\ \mu s$) equilibrium simulations in *cis* and *trans* state of photoswitch.

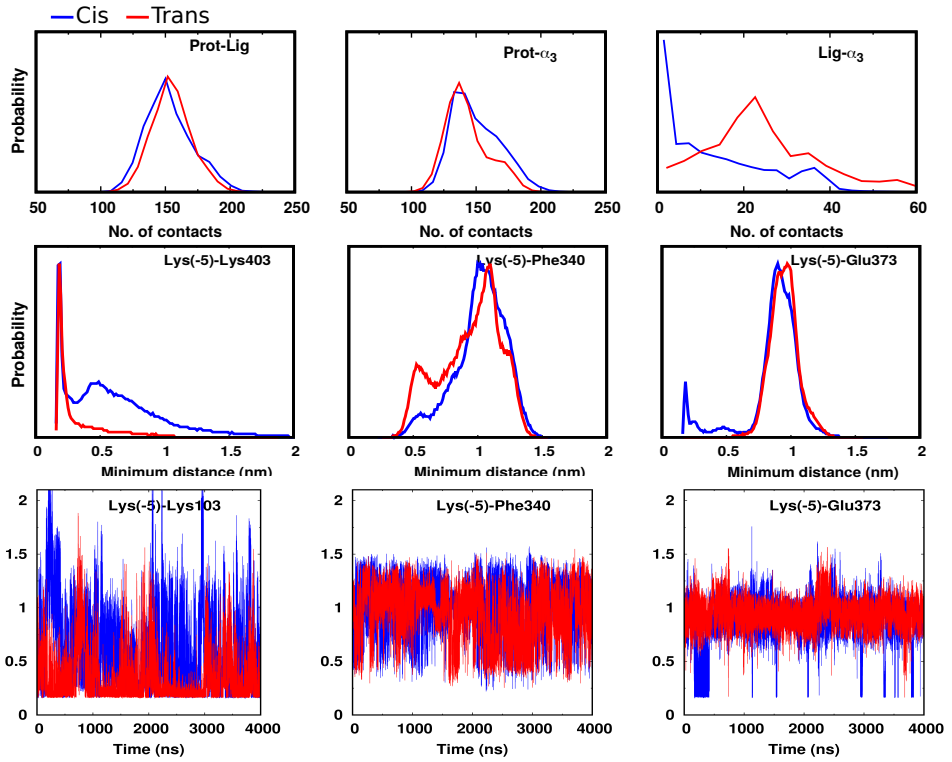


FIGURE 5.7: (Top) Probability of number of contacts among protein, ligand and α_3. (Middle) Contacts formation of Lys(-5) with residues of α_3, β_3 and α_2 from *cis* and *trans* equilibrium MD simulations. (Bottom) Time evolution of contacts.

The equilibrium MD simulations reveal a significant increase in total number of contacts between ligand and α_3 in *trans* state (Fig.5.7). Furthermore, contact analysis shows that Lys(-5) of ligand develops a strong contact with Lys403 of α_3 (Fig.5.8). Accompanying nonequilibrium contact analysis further reveals that this contact achieves stability at around $\approx$ 100 ns. In addition, *cis-to-trans* photoswitching appears to enforce the center of mass of ligand away from binding pocket by 0.2 Å (figure 5.9).
We propose that center of mass motion of ligand away from binding pocket at around $\approx$ 50 – 300 ns by 0.2 Å represents the dynamical allosteric response whereas the formation of Lys(-5) – Lys403 contact around the similar times represents the structural allosteric response. Therefore, we conclude that timeline of $\approx$ 50 – 300 ns is sufficient to resolve

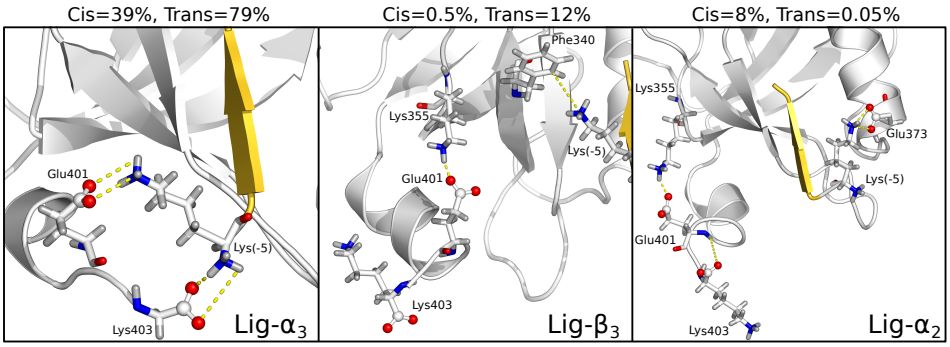


FIGURE 5.8: A representative cartoon showing the orientational preference of Lys(-5) towards α_3, β_3 and α_2. Respective contact survival percentage in *cis* and *trans* state of photoswitch is displayed at the top of each conformation.

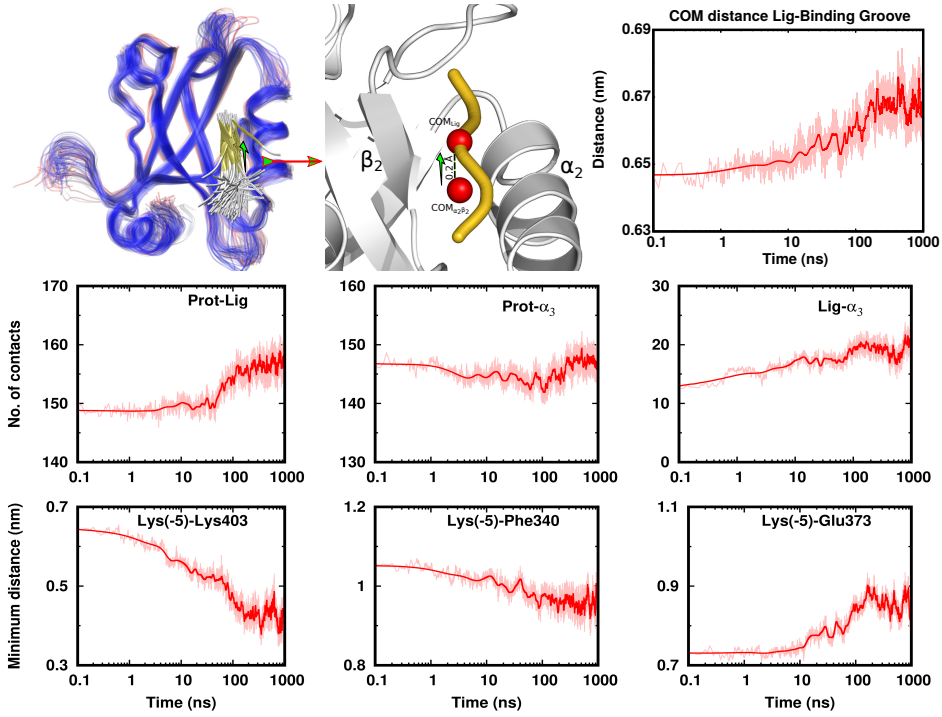


FIGURE 5.9: (Top) Ligand motion out of binding pocket and respective distance between center of mass of ligand and center of binding pocket of nonequilibrium ensemble. (Middle) Time evolution of number of contacts among protein, ligand and α_3. (Bottom) Contacts time evolution of Lys(-5) with residues of α_3, β_3 and α_2.

first phase of overall relaxation and allosteric response of ligand upon *cis to trans* photoswitching of distant α_3 in PDZ3 domain. Based on above observations, we argue that reduction in $\approx$ 25–folds binding affinity observed in intrinsic tryptophan fluorescence and ITC experiments is due to coupled dynamical and structural adjustments. Further, experimental and computational effort is underway to investigate the validity of this proposal.

Chapter 6

Conclusions and Outlook

Energy and signal transport together encompass protein dynamics on multiple timescales ranging from picoseconds to microseconds. Overall, we have demonstrated that nonequilibrium MD simulations are a reliable approach to gain insight into the dynamics on atomic scales. Theoretical predictions of nonequilibrium simulations can be directly compared with time-resolved experiments. In a joint theoretical-experimental approach, we have thoroughly described the mechanisms of vibrational energy transport, protein-peptide interactions, and allostery in proteins. Furthermore, we have shown that the Aha label has the ability to report on minuscule changes in structural dynamics around its environment.

Starting from vibrational energy transport on picosecond timescales, we have developed a general simulation strategy that mimics the experimental procedure of Bredenbeck and coworkers.[37,82] By using this procedure, nonequilibrium simulations quantitatively reproduce the experimental cooling times ($\approx$ 6-9 ps) at room temperature[44,82] and reveal energy transport pathways in TrpZip2 and PDZ2. Due to their efficiency, our nonequilibrium simulations may serve as a reference for more approximate calculations of the energy flow. Furthermore, by using these simulations we have determined the parameters of the master equation to model energy transport in proteins, we provided two general scaling rules (eq: 2.5 and 2.10) for backbone energy transport and contact energy transport. The relations require only respective diffusion coefficients ($D_\mathrm{B} = 1.1\,\mathrm{nm}^2/\mathrm{ps}$ and $D_\mathrm{C} = 2.1 \cdot 10^{-3}\,\mathrm{nm}^2/\mathrm{ps}$) and interatom distances along the backbone or mean square distance of contacting atoms respectively. In addition, we have identified the relevant energy transport pathways of the considered systems. A comparison with experiments reveals that MD calculations overestimate the peak times of energy transport by about a factor 2 due to their classical nature. However, the global scaling by a factor 3.1 of master equation parameters (heater rate k_h, the backbone diffusion coefficient D_B, the contact transport coefficient D_C and heater contact rates) reproduces the experimental results. Since many studies link energy transport pathways with the pathways of allosteric mechanism, the combination of experimental results and master equation models of an allosteric protein such as PDZ3 may provide answer to the long-standing question about the existence of common pathways for energy transport and allostric communication. While theoretical results are already available, the experimental work on the energy transport of PDZ3 domain is currently underway.

Furthermore, we carried out an in-depth experimental and computational analysis of the Aha label as a sensitive IR probe. While 2D IR spectroscopy of AHA-labelled PDZ2 domain reveals a distinct response of one of the mutant (L78AHA), we found that instead of a shift in frequency, the Aha label in this case reveals a change in intensity. Accompanying quantum mechanical calculations show that this response originates from an electronic interaction between the azido group of Aha and the sulfonate group of the azobenzene moiety. This constitutes a new mode of spectral response of Aha label that reports on electronic instead of an electrostatic interaction of azido groups with the protein/water surrounding. Moreover, by incorporating Aha labels inside the protein domain and measuring the changes by excluding and including a peptide ligand, we have demonstrated that 2D IR spectroscopy can reliably measure frequency shifts of the Aha label as small as $\approx$1 cm^{-1}. Vibartional frequency shifts from MD data qualitatively reproduce this effect, thus providing an atomistic picture of the interaction of the Aha label with side chains of Glu(-5) that causes the frequency shift. Present studies demonstrate that Aha can be employed as a specific IR reporter not only for big changes of its chemical environment (e.g., protein folding/unfolding), but also for very subtle changes of the electronic/electrostatic environment at the protein surface.

In Chapter 4, by employing Rnase S as a model system, we have designed five photoswitchable S-peptides and demonstrated that the helical content of S-peptide determines its binding affinity with S-protein. We have shown that helical content can be modulated in a very controlled manner and hence the protein-peptide interactions. MD simulations accompanied with a simple Quantitative Structure-Activity Relationship ansatz have shown that the α-helicity is inversely related to the experimental K_d's, suggesting that relatively short (sub-μs) MD simulations are indicative of the long-time binding or dissociation behavior of the RNase S complex. Most remarkable, by applying this strategy we have observed that one of the photoswitchable S-peptides (S-pep(6,13)) is no more bound to the S-protein. Taking advantage of the photocontrolled unbinding of this system, the long-standing question whether the process can be described as "conformational selection" or as "induced fit" can be addressed by again employing a combination of nonequilibrium simulations and transient IR experiments.

By combining transient IR spectroscopy, extensive non-equilibrium MD simulations, and Markov state modeling, in chapter 5, we have described the ligand-induced conformational transitions in the PDZ2 domain that are thought to be responsible for the allosteric communication within the proteins. The joint experimental-theoretical study shows that reversible *trans*-to-*cis* ligand-switching causes a conformational transition of PDZ2 in terms of the mean structure as well as changes the protein's fluctuations. A few structurally well defined states and dynamics distributed over four decades starting from a nanosecond to $\sim$ 10 microseconds timescale fully characterize the response of allosteric transitions. To further investigate the mechanisms of allostery, computational and experimental studies of α_3-switched PDZ3 domain are underway.

List of Figures

List of Tables

Bibliography

[1] Bruce Alberts et al. *Essentials of Cell Biology*. New York: W.W. Norton and Company, 2013, pp. 1–50.

[2] Bruce Alberts et al. *Essentials of Cell Biology*. New York: W.W. Norton and Company, 2013, pp. 171–250.

[3] Stanfield RL and Wilson IA. "Protein-peptide interactions". In: *Curr. Opin. Struct. Biol.* 5 (1995), pp. 103–113.

[4] O'Connor C. M. and Adams J. U. *Essentials of Cell Biology*. Cambridge, MA: NPG Education: Nature Scitable, 2010.

[5] J. M. Berg, J. L. Tymoczko, and L. Stryer. *Biochemistry*. 41 Madison Ave., New York: W.H. Freeman and Co, 2001.

[6] John Boyle. "Lehninger principles of biochemistry (4th ed.): Nelson, D. and Cox, M." In: *Biochemistry and Molecular Biology Education* 33.1 (2005), pp. 74–75.

[7] Bruce Alberts et al. *Essentials of Cell Biology*. New York: W.W. Norton and Company, 2013, pp. 121–170.

[8] Dorothee Kern and Erik RP Zuiderweg. "The role of dynamics in allosteric regulation". In: *Current Opinion in Structural Biology* 13.6 (2003), pp. 748 –757.

[9] J.M. Yon, D. Perahia, and C. Ghélis. "Conformational dynamics and enzyme activity". In: *Biochimie* 80.1 (1998), pp. 33 –42.

[10] Mouzhe Xie et al. "Functional protein dynamics on uncharted time scales detected by nanoparticle-assisted NMR spin relaxation". In: 5.8 (2019). DOI: 10.1126/sciadv.aax5560.

[11] Fritz Parak. "Proteins in action: The physics of structural fluctuations and conformational changes". In: *Current opinion in structural biology* 13 (Oct. 2003), pp. 552–7.

[12] Chong Fang and Longteng Tang. "Mapping Structural Dynamics of Proteins with Femtosecond Stimulated Raman Spectroscopy". In: *Annual Review of Physical Chemistry* 71.1 (2020), pp. 239–265.

[13] Katherine Henzler-Wildman and Dorothee Kern. "Dynamic personalities of proteins". In: *Nature* 450 (2007), pp. 964–972.

[14] G. Stock and P. Hamm. "A nonequilibrium approach to allosteric communication". In: *Phil. Trans. B* 373 (2018), p. 20170187.

[15] S. Buchenberg, F. Sittel, and G. Stock. "Time-resolved observation of protein allosteric communication". In: *Proc. Natl. Acad. Sci. USA* 114 (2017), E6804.

[16] B. Buchli et al. "Kinetic response of a photoperturbed allosteric protein". In: *Proc. Natl. Acad. Sci. USA* 110 (2013), p. 11725.

[17] H. Nakhaei Motlagh et al. "The ensemble nature of allostery". In: *Nature* 508 (2014), pp. 331–339.

[18] Wodak et al. "Allostery in Its Many Disguises: From Theory to Applications". In: *Structure* 27 (2019), pp. 566–578.

[19] J. Monod, J. Wyman, and J. P. Changeux. "ON THE NATURE OF ALLOSTERIC TRANSITIONS - A PLAUSIBLE MODEL". In: *J. Mol. Biol.* 12 (1965), pp. 88–118.

[20] Jean-Pierre Changeux and Stuart J. Edelstein. "Allosteric Mechanisms of Signal Transduction". In: *Science* 308.5727 (2005), pp. 1424–1428. DOI: 10.1126/science.1108595.

[21] D. E. Koshland, G. Némethy, and D. Filmer. "Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits*". In: *Biochemistry* 5.1 (1966), pp. 365–385.

[22] Peter Csermely, Robin Palotai, and Ruth Nussinov. "Induced fit, conformational selection and independent dynamic segments: an extended view of binding events". In: *Trends in Biochemical Sciences* 35.10 (2010), pp. 539 –546. ISSN: 0968-0004.

[23] Q. Cui and M. Karplus. "Allostery and cooperativity revisited". In: *Protein Science* 17 (2008), pp. 1295–1307.

[24] I Bahar, C Chennubhotla, and D. Tobi. "Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation". In: *Curr. Opin. Struct. Biol* 17 (2007), pp. 633–640.

[25] Brian F. Volkman et al. "Two-State Allosteric Behavior in a Single-Domain Signaling Protein". In: *Science* 291.5512 (2001), pp. 2429–2433. DOI: 10.1126/science.291.5512.2429.

[26] A Cooper and D T F Dryden. "Allostery without conformational change". In: *Eur. Biophys. J.* 11 (1984), pp. 103–109.

[27] J. Guo and H.-X. Zhou. "Protein allostery and conformational dynamics". In: *Chem. Rev.* 116 (2016), p. 6503.

[28] O. Schueler-Furman and S. J. Wodak. "Computational approaches to investigating allostery". In: *Curr. Op. Struct. Bio.* 41 (2016).

[29] A. Dhulesia, J. Gsponer, and M Vendruscolo. "Mapping of Two Networks of Residues that Exhibit Structural and Dynamical Changes upon Binding in a PDZ Domain Protein." In: *J. Am. Chem. Soc.* 130 (2008), pp. 8931–8939.

[30] S. W. Lockless and R Ranganathan. "Evolutionarily Conserved pathways of energetic connectivity in protein families". In: *Science* 286 (1999), pp. 295–299.

[31] T. Ishikura and T. Yamato. "Energy transfer pathways relevant for long-range intramolecular signaling of photosensory protein revealed by microscopic energy conductivity analysis". In: *Chem. Phys. Lett.* 432 (2006), p. 533.

[32] D. M. Leitner and J. E. Straub. *Proteins: Energy, Heat and Signal Flow*. Taylor and Francis/CRC Press, London, 2009.

[33] L. Maggi, P. Carloni, and G. Rossetti. "Vibrational energy in proteins correlates with topology". In: *J. Phys. Chem. Lett.* 9 (2018), p. 6393.

[34] Hamm Lab. URL: https://www.chem.uzh.ch/en/research/groups/hamm/research.html.

[35] Bredenbeck Group. URL: http://www.biophys.uni-frankfurt.de/~bredenbeck/.

[36] G. Stock, J. Wachtveitl, and H. Grubmueller. "Nonequilibrium dynamics in biomolecules". In: *Chemical Physics* 323.1 (2006). Nonequilibrium Dynamics in Biomolecules, p. 1. DOI: https://doi.org/10.1016/j.chemphys.2005.08.043.

[37] H. M. Muller-Werkmeister and J. Bredenbeck. "A donor-acceptor pair for the real time study of vibrational energy transfer in proteins". In: *Phys. Chem. Chem. Phys.* 16 (2014), p. 3261.

[38] P.H. Nguyen, S.M. Park, and G.Stock. "Nonequilibrium molecular dynamics simulation of the energy transport through a peptide helix". In: *J. Chem. Phys.* 132 (2010), p. 025102.

[39] E. Deniz et al. "Energy flow pathways in Proteins: Competition between backbone and contact transport". In: *Preparation* (2020).

[40] P H Nguyen and G Stock. "Nonequilibrium molecular dynamics simulation of a photoswitchable peptide". In: *Chem. Phys.* 323 (2006), pp. 36–44.

[41] V. Botan et al. "Energy transport in peptide helices". In: *Proc. Natl. Acad. Sci. USA* 104 (2007), p. 12749.

[42] E. Backus et al. "Energy transport in peptide helices: A comparison between UV IR pumping". In: *J. Phys. Chem. B* 112 (2008), p. 9091.

[43] E. Backus et al. "Structural flexibility of a helical peptide regulates vibrational energy transport properties". In: *J. Phys. Chem. B* 112 (2008), p. 15487.

[44] J. Bredenbeck et al. "α-helix formation in a photoswitchable peptide tracked from picoseconds to microseconds by time-resolved IR spectroscopy". In: *Proc. Natl. Acad. Sci. USA* 102 (2005), p. 2379.

[45] H. J. C. Berendsen. *Simulating the Physical World*. Cambridge: Cambridge University Press, 2007.

[46] J. M. Haile. *Molecular Dynamics Simulation: Elementary Methods*. 1st. USA: John Wiley & Sons, Inc., 1992. ISBN: 0471819662.

[47] D. C. Rapaport. *The Art of Molecular Dynamics Simulation*. 2nd ed. Cambridge University Press, 2004. DOI: 10.1017/CBO9780511816581.

[48] M. J. Abraham et al. "Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". In: *SoftwareX* 1–2 (2015), p. 19.

[49] M. Yokokawa et al. "The K computer: Japanese next-generation supercomputer development project". In: (2011), pp. 371–372.

[50] David E. Shaw et al. "Millisecond-Scale Molecular Dynamics Simulations on Anton". In: *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*. SC '09. Portland, Oregon: Association for Computing Machinery, 2009.

[51] Ian Jolliffe. "Principal Component Analysis". In: *International Encyclopedia of Statistical Science*. Ed. by Miodrag Lovric. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 1094–1096. DOI: 10.1007/978-3-642-04898-2_455.

[52] M. Ernst, F. Sittel, and G. Stock. "Contact- and distance-based principal component analysis of protein dynamics". In: *J. Chem. Phys.* 143 (2015), p. 244114.

[53] F. Sittel and G. Stock. "Perspective: Identification of collective variables and metastable states of protein dynamics". In: *J. Chem. Phys.* 149 (2018), p. 150901.

[54] F. Sittel and G. Robust density-based clustering to identify metastable conformational states of proteins Stock. In: *J. Chem. Theory Comp.* 12 (2016), pp. 2426–2435.

[55] Daniel Nagel et al. "Dynamical coring of Markov state models". In: *J. Chem. Phys.* 150.9 (2019), p. 094111.

[56] Frank Noé et al. "Dynamical fingerprints for probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments". In: *Proc. Natl. Acad. Sci. U. S. A.* 108.12 (2011), pp. 4822–4827.

[57] Frank Jensen. *Introduction to Computational Chemistry*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 1999.

[58] G. Craig Hill. *Molecular Mechanics across Chemistry (Rappé, Anthony K. and Casewit, Carla J.)* Vol. 76. 1. 1999, p. 31.

[59] Wendy D. Cornell et al. "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids and Organic Molecules". In: *Journal of the American Chemical Society* 117.19 (1995), pp. 5179–5197.

[60] W. J. Hehre. *A Guide to Molecular Mechanics and Quantum Chemical Calculations, Wavefunction*. 2003, pp. 436–441.

[61] Loup Verlet. "Computer Experiments on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules". In: *Phys. Rev.* 159 (1 1967), pp. 98–103.

[62] William 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: *The Journal of Chemical Physics* 76.1 (1982), pp. 637–649.

[63] R.W Hockney, S.P Goel, and J.W Eastwood. "Quiet high-resolution computer models of a plasma". In: *Journal of Computational Physics* 14.2 (1974), pp. 148 – 158.

[64] Rainer Hegger et al. "How Complex is the Dynamics of Peptide Folding?" In: 98 (2007), p. 028102.

[65] Stefano Piana and Alessandro Laio. "Advillin Folding Takes Place on a Hypersurface of Small Dimensionality". In: *Phys. Rev. Lett.* 101 (2008), p. 208101.

[66] John D Chodera and Frank Noé. "Markov state models of biomolecular conformational dynamics". In: *Current Opinion in Structural Biology* 25 (2014), pp. 135 –144.

[67] J. R. Norris. *Markov Chains*. Cambridge University Press, 1997.

[68] Gregory R. Bowman, Vijay S. Pande, and Frank Noe. *An Introduction to Markov State Models*. Heidelberg: Springer, 2013.

[69] Martin K. Scherer et al. "PyEMMA 2: A Software Package for Estimation, Validation and Analysis of Markov Models". In: *Journal of Chemical Theory and Computation* 11.11 (2015), pp. 5525–5542.

[70] J.-H. Choi, K.-I. Oh, and M Cho. "Azido-derivatized compounds as IR probes of local electrostatic environment: Theoretical studies". In: *J. Chem. Phys.* 129 (2008), p. 174512.

[71] J. H. Choi, D. Raleigh, and M Cho. "Azido homoalanine is a useful infrared probe for monitoring local electrostatistics and side-chain solvation in proteins". In: *J. Phys. Chem. Lett.* 2 (2011), pp. 2158–2162.

[72] G. Cerullo et al. "Photosynthetic Light Harvesting by Carotenoids: Detection of an Intermediate Excited State". In: *Science* 298.5602 (2002), pp. 2395–2398.

[73] E. R. Henry, W. A. Eaton, and R. M. Hochstrasser. "Molecular-dynamics simulations of cooling in laser-excited heme-proteins". In: *Proc. Natl. Acad. Sci. USA* 83 (1986), p. 8982.

[74] Y. Mizutani and T. Kitagawa. "Direct observation of cooling of heme upon photodissociation of carbonmonoxy myoglobin". In: *Science* 278 (1997), p. 443.

[75] N. Ota and D. A. Agard. "Intramolecular signaling pathways revealed by molecular anisotropic thermal diffusion". In: *J. Mol. Biol.* 351 (2005), p. 345.

[76] P. H. Nguyen, P. Derreumaux, and G. Stock. "Energy flow and long-range correlations in guanine-binding riboswitch: A nonequilibrium molecular dynamics study". In: *J. Phys. Chem. B* 113 (2009), p. 9340.

[77] L. Martinez et al. "Mapping the intramolecular vibrational energy flow in proteins reveals functionally important residues". In: *J. Phys. Chem. Lett.* 2 (2011), p. 2073.

[78] J. K. Weber, D. Shukla, and V. S. Pande. "Heat dissipation guides activation in signaling proteins". In: *Proc. Natl. Acad. Sci. USA* 112 (2015), p. 10377.

[79] Natalia I. Rubtsova and Igor V. Rubtsov. "Vibrational Energy Transport in Molecules Studied by Relaxation-Assisted Two-Dimensional Infrared Spectroscopy". In: *Annual Review of Physical Chemistry* 66.1 (2015), pp. 717–738.

[80] N. Fujii et al. "Observing vibrational energy flow in a protein with the spatial resolution of a single amino acid residue". In: *J. Phys. Chem. Lett.* 5 (2014), p. 3269.

[81] S. Yamashita et al. "Vibrational energy transfer from heme through atomic contacts in proteins". In: *J. Phys. Chem. B* 122 (2018), p. 5877.

[82] T. Baumann et al. "Site-resolved observation of vibrational energy transfer using a genetically encoded ultrafast heater". In: *Angew. Chem. Int. Ed.* 58 (2019), p. 2899.

[83] A. Gulzar et al. "Energy transport pathways in proteins: A nonequilibrium molecular dynamics simulation study". In: *J. Chem. Theory Comput.* 15 (2019), p. 5750.

[84] A. G. Cochran, N. J. Skelton, and M. A. Starovasnik. "Tryptophan zippers: Stable, monomeric beta-hairpins". In: *Proc. Natl. Acad. Sci. USA* 98 (2001), p. 5578.

[85] D. M. Leitner et al. "Vibrational energy flow in the villin headpiece subdomain: Master equation simulations". In: *J. Chem. Phys.* 142 (2015), p. 075101.

[86] S. Buchenberg, D. M. Leitner, and G. Stock. "Scaling rules for vibrational energy transport in globular proteins". In: *J. Phys. Chem. Lett.* 7 (2016), p. 25.

[87] P. Hamm, S. M. Ohline, and W. Zinth. "Vibrational cooling after ultrafast photoisomerization of azobenzene measured by femtosecond infrared spectroscopy". In: *J. Chem. Phys.* 106 (1997), p. 519.

[88] Waegele et al. "Site-Specific Spectroscopic Reporters of the Local Electric Field, Hydration, Structure and Dynamics of Biomolecules." In: *J. Phys. Chem. Lett.* 2 (2011), pp. 2598–2609.

[89] R. Adhikary et al. "IR Probes of Protein Microenvironments: Utility and Potential for Perturbation". In: *ChemPhysChem* 5 (2014), pp. 849–853.

[90] H. Kim and M Cho. "Infrared probes for studying the structure and dynamics of biomolecules". In: *Chem. Rev.* 113 (2013), pp. 5817–5847.

[91] K. L. Koziol et al. "Fast Infrared Spectroscopy of Protein Dnamics: Advancing Sensitivity and Selectivity". In: *Curr. Opin. Struct. Biol.* 34 (2015), pp. 1–6.

[92] J. Ma et al. "Site-Specific Infrared Probes of Proteins". In: *Annu. Rev. Phys. Chem.* 66 (2015), pp. 357–377.

[93] R. Bloem et al. "Ligand binding studied by 2D IR spectroscopy using the azidohomoalanine label". In: *J. Phys. Chem. B* 116 (2012), pp. 13705–13712.

[94] P. J. M. Johnson, K. L. Koziol, and P Hamm. "Quantifying Biomolecular Recognition with Site-Specific 2D Infrared Probes". In: *J. Phys. Chem. Lett.* 8 (2017), pp. 2280–2284.

[95] N. Blochliger, M. Xu, and A Caflisch. "Peptide binding to a PDZ domain by electrostatic steering via nonnative salt bridges". In: *Biophys. J.* 108 (2015), pp. 2362–2370.

[96] Z Huang. "The chemical biology of apoptosis: Exploring protein-protein interactions and the life and death of cells with small molecules". In: *Chem. Biol.* 9 (2002), pp. 1059–1072.

[97] A. A. Beharry and G. A Woolley. "Azobenzene photoswitches for biomolecules". In: *Chem. Soc. Rev.* 40 (2011), pp. 4422–4437.

[98] T. S. Zatsepin et al. "Design of photocontrolled biomolecules based on azobenzene derivatives". In: *Russ. Chem. Rev.* 82 (2013), pp. 942–963.

[99] D. G. Flint et al. "Using an Azobenzene Cross-Linker to Either Increase or Decrease Peptide Helix Content upon Trans-to-Cis Photoisomerization". In: *Chem. Biol.* 9 (2002), pp. 391–397.

[100] F. M. Richards and P. J Vithayathil. "The Preparation of Subtilisin-modified Ribonuclease and the Separation of the Peptide and Protein Components". In: *J. Biol. Chem.* 234 (1959), pp. 1459–1465.

[101] A. Wlodawer and L Sjolin. "Hydrogen exchange in RNase A: Neutron diffraction study". In: *Proc. Natl Acad. Sci. USA* 79 (1982), pp. 1418–1422.

[102] A. A. Schreier and R. L Baldwin. "Mechanism of Dissociation of S-Peptide from Ribonuclease". In: *Biochemistry* 16 (1976), pp. 4203–4209.

[103] J. M. Goldberg and R. L Baldwin. “A specific transition state for S-peptide combining with folded S-protein and then refolding”. In: *Proc. Natl. Acad. Sci. USA* 96 (1999), pp. 2019–2024.

[104] A. Bachmann et al. “Mapping backbone and side-chain interactions in the transition state of a coupled protein folding and binding reaction”. In: *Proc. Nat. Acad. Sci. USA* 108 (2011), pp. 3952–3957.

[105] J. Dogan, S. Gianni, and P Jemth. “The binding mechanisms of intrinsically disordered proteins”. In: *Phys. Chem. Chem. Phys.* 16 (2014), pp. 6323–6331.

[106] M. P. Luitz, R. Bomblies, and M Zacharias. “Comparative Molecular DynamicsAnalysis of RNase-S Complex Formation”. In: *Biophys. J.* 113 (2017), pp. 1466–1474.

[107] J. R. Kumita, O. S. Smart, and G. A Woolley. “Photo-control of helix content in a short peptide”. In: *Proc. Natl. Acd. Sci. USA* 97 (2000), pp. 3803–3808.

[108] G. A Woolley. “Photocontrolling peptide alpha helices”. In: *Acc. Chem. Res.* 38 (2005), pp. 486–493.

[109] J. A. Ihalainen et al. “Alpha-Helix folding in the presence of structural constraints.” In: *Proc. Natl. Acad. Sci. USA* 105 (2008), pp. 9588–9593.

[110] Jean-Pierre Changeux. “Allostery and the Monod-Wyman-Changeux Model After 50 Years”. In: *Annu. Rev. Biophys.* 41.1 (2012), pp. 103–133.

[111] E. J. Fuentes, C. J. Der, and A. L Lee. “Ligand-dependent dynamics and intramolecular signaling in a PDZ Domain”. In: *J. Mol. Biol.* 335 (2004), pp. 1105–1115.

[112] E. J. Fuentes et al. “Evaluation of Energetic and Dynamic Coupling Networks in a PDZ Domain Protein.” In: *J.Mol. Biol.* 364 (2006), pp. 337–351.

[113] C. M. Petit et al. “Hidden dynamic allostery in a PDZ domain”. In: *Proc. Natl. Acad. Sci. USA* 106 (2009), p. 18249.

[114] G. Kozlov et al. “Solution structure of the PDZ2 domain from cytosolic human phosphatase hPTP1E complexed with a peptide reveals contribution of the beta 2-beta 3 loop to PDZ domain-ligand interactions”. In: *J. Mol. Biol.* 320 (2002), pp. 813–820.

[115] Z. Zhang et al. “A Water-Soluble Azobenzene Cross-Linker for Photocontrol of Peptide Conformation”. In: *Bioconjugate Chem.* 14 (2003), pp. 824–829.

[116] Hans Frauenfelder, Stephen G Sligar, and Peter G Wolynes. “The Energy Landscapes and Motions of Proteins”. In: *Science* 254 (1991), pp. 1598–1603.

[117] Ken A. Dill and H. S. Chan. “From Levinthal to Pathways to Funnels: The "New View" of Protein Folding Kinetics”. In: *Nat. Struct. Bio.* 4 (1997), pp. 10–19.

[118] Robert G. Smock and Lila M. Gierasch. “Sending signals dynamically”. In: *Science* 324.5924 (2009), pp. 198–203.

[119] Chung Jung Tsai and Ruth Nussinov. “A Unified View of "How Allostery Works"”. In: *PLoS Comput. Biol.* 10.2 (2014).

[120] Vincent J. Hilser, James O. Wrabl, and Hesam N. Motlagh. "Structural and Energetic Basis of Allostery". In: *Annual Review of Biophysics* 41.1 (2012), pp. 585–609.

[121] D. A. Doyle et al. "Crystal structures of a complexed and peptide-free membrane protein–binding domain: Molecular basis of peptide recognition by PDZ". In: *Cell* 85 (1996), p. 1067.

[122] Morais Cabral JH, Petosa C, and et al. Sutcliffe MJ. "Crystal structure of a PDZ domain". In: *Nature* 382 (1996), 649-–652.

[123] J. Deak et al. "Vibrational energy transfer across a reverse micelle surfactant layer". In: *Science* 306 (2004), p. 473.

[124] G. Li, D. Magana, and R. B. Dyer. "Anisotropic energy flow and allosteric ligand binding in albumin". In: *Nat. Commun.* 5 (2014), p. 3100.

[125] I. V. Rubtsov and A. L. Burin. "Perspective: Ballistic and diffusive vibrational energy transport in molecules". In: *J. Chem. Phys.* 150 (2019), p. 020901.

[126] for the sake of a simple physical picture we invoke the kinetic energy of an atom group as a measure of local temperature Although the concept of a local temperature is difficult to define in a strict sense. In: ().

[127] H.-J. Lee and J. J. Zheng. "PDZ domains and their binding partners: structure, specificity and modification". In: *Cell Commun. Signal.* 8 (2010).

[128] Y. Kong and M. Karplus. "Signaling pathways of PDZ2 domain: A molecular dynamics interaction correlation analysis". In: *Proteins* 74 (2009), p. 145.

[129] T. Ishikura et al. "Energy exchange network of inter-residue interactions within a thermally fluctuating protein molecule: A computational study". In: *J. Comput. Chem.* 36 (2015), p. 1709.

[130] W. B. Wang et al. "Energy transport pathway in proteins: Insights from nonequilibrium molecular dynamics with elastic network model". In: *Sci. Rep.* 8 (2018), p. 9487.

[131] S.M. Park, P.H. Nguyen, and G.Stock. "Molecular dynamics simulation of cooling: Heat transfer from a photoexcited peptide to the solvent". In: *J. Chem. Phys.* 131 (2009), p. 184503.

[132] P. H. Nguyen and G. Stock. "Nonequilibrium molecular-dynamics study of the vibrational energy relaxation of peptides in water". In: *J. Chem. Phys.* 119 (2003), p. 11350.

[133] H. Fujisaki and J. E. Straub. "Vibrational energy relaxation in proteins". In: *Proc. Natl. Acad. Sci. USA* 102 (2005), p. 6726.

[134] M. Angel Soler et al. "A method for analyzing the vibrational energy flow in biomolecules in solution". In: *J. Chem. Phys.* 135 (2011).

[135] X. Yu and D. Leitner. "Vibrational energy transfer and heat conduction in a protein". In: *J. Phys. Chem. B* 107 (2003), p. 1698.

[136] D. M. Leitner. "Frequency-resolved communication maps for proteins and other nanoscale materials". In: *J. Chem. Phys.* 130 (2009), p. 195101.

[137] A. A. S. T. Ribeiro and V. Ortiz. "Determination of signaling pathways in proteins through network theory: Importance of the topology". In: *J. Chem. Theory Comput.* 10 (2014), p. 1762.

[138] K. Reid, T. Yamato, and D. M. Leitner. "Scaling of rates of vibrational energy transfer in proteins with equilibrium dynamics and entropy". In: *J. Phys. Chem. B* 122 (2018), p. 9331.

[139] V. Hornak et al. "Comparison of multiple Amber force fields and development of improved protein backbone parameters". In: *Proteins* 65 (2006), p. 712.

[140] R. B. Best and G. Hummer. "Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides". In: *J. Phys. Chem. B* 113 (2009), p. 9004.

[141] K. Lindorff-Larsen et al. "Improved side-chain torsion potentials for the Amber ff99sb protein force field". In: *Proteins* 78 (2010), p. 1950.

[142] J. Wang et al. "Antechamber, an accessory software package for molecular mechanical calculations". In: *J. Comput. Chem.* 25 (2005), p. 1157.

[143] M. J. et al Frisch. "Gaussian 09". In: *Gaussian Inc. Wallingford CT, 2009.* ().

[144] B. Stucki-Buchli et al. "2D-IR spectroscopy of an AHA labelled photoswitchable PDZ2 domain". In: *J. Phys. Chem. A* 121 (2017), pp. 9435 –9445.

[145] T. Darden, D. York, and L. Petersen. "Particle mesh Ewald: An N log(N) method for Ewald sums in large systems". In: *J. Chem. Phys.* 98 (1993), p. 10089.

[146] G. Bussi et al. "Canonical sampling through velocity rescaling". In: *J. Chem. Phys.* 126 (2007), p. 014101.

[147] H. J. C. Berendsen et al. "Molecular dynamics with coupling to an external bath". In: *J. Chem. Phys.* 81 (1984), p. 3684.

[148] W. L. Jorgensen et al. "Comparison of simple potential functions for simulating liquid water". In: *J. Chem. Phys.* 79 (1983), p. 926.

[149] A. J. Wurzer et al. "Comprehensive measurement of the S1 azulene relaxation dynamics and observation of vibrational wavepacket motion". In: *Chem. Phys. Lett.* 299 (1999), p. 296.

[150] D. Schwarzer et al. "Vibrational energy transfer in highly excited bridged azulene-aryl compounds: Direct observation of energy flow through aliphatic chains and into the solvent". In: *J. Phys. Chem. A* 106 (2002), p. 8019.

[151] J. Gans and D. Shalloway. "Shadow mass and the relationship between velocity and momentum in symplectic numerical integration". In: *Phys. Rev. E* 61 (2000), p. 4587.

[152] J. Helbing et al. "Temperature dependence of the heat diffusivity of proteins". In: *J. Phys. Chem. B* 116 (2012), p. 2620.

[153] B. J. Ka and E. Geva. "Classical vs quantum vibrational energy relaxation pathways in solvated polyatomic molecules". In: *J. Phys. Chem. A* 110 (2006), p. 13131.

[154] G. Stock. "Classical simulation of quantum energy flow in biomolecules". In: *Phys. Rev. Lett.* 102 (2009), p. 118301.

[155] F. Lehner et al. "Impact of azidohomoalanine incorporation on protein structure and ligand binding". In: *ChemBioChem* 18 (2017), p. 2340.

[156] N. V. Dokholyan. "Controlling allosteric networks in proteins". In: *Chem. Rev.* 116 (2016), p. 6463.

[157] L. Cheng, V. Knecht, and G. Stock. "Long-range conformational response of a PDZ domain to ligand binding and release: A molecular dynamics study". In: *J. Chem. Theory Comput.* 12 (2016), p. 870.

[158] Christopher I. Bayly et al. "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model". In: *The Journal of Physical Chemistry* 97.40 (1993), pp. 10269–10280.

[159] C. Zanobini et al. "Azidohomoalanine: A minimally-invasive, versatile and sensitive infrared label in proteins to study ligand binding". In: *J. Phys. Chem. B* 122 (2018), p. 10118.

[160] W. J. Schreier et al. "Following the energy transfer in and out of a polyproline peptide". In: *Pep. Sci.* 100 (2013), p. 38.

[161] Frank Noé et al. "Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations". In: *Proceedings of the National Academy of Sciences* 106.45 (2009), pp. 19011–19016.

[162] Luis Valiño Borau. "Master equation model of energy transport in proteins: From Classical dynamics to Quantum corrections".

[163] W. Stacklies, F. Xia, and F. Grater. "Dynamic allostery in the methionine repressor revealed by force distribution analysis". In: *PLoS Comput. Biol.* 5 (2009), e1000574, 11.

[164] S. A. Egorov, K. F. Everitt, and J. L. Skinner. "Quantum Dynamics and Vibrational Relaxation". In: *J. Phys. Chem. A* 103 (1999), p. 9494.

[165] M. Koziński, S. Garrett-Roe, and P Hamm. "2D-IR Spectroscopy of the Sulfhydryl Band of Cysteines in the Hydrophobic Core of Proteins". In: *J. Phys. Chem. B* 112 (2008), pp. 7645–7650.

[166] S. R. G. Naraharisetty et al. "C-D Modes of Deuterated Side Chain of Leucine as Structural Reporters via Dual-frequency Two-dimensional Infrared Spectroscopy". In: *J. Phys. Chem. B* 113 (2009), pp. 4940–4946.

[167] M. Schade et al. "Vibrational Energy Transport in Peptide Helices after Excitation of C-D Modes in Leu-D10". In: *J. Phys. Chem. B* 113 (2009), pp. 13393–13397.

[168] J. Zimmermann et al. "Carbon-Deuterium Bonds as Site-Specific and Nonperturbative Probes for Time-Resolved Studies of Protein Dynamics and Folding". In: *J. Phys. Chem. Lett.* 2 (2011), pp. 412–416.

[169] P. A. Anfinrud, C. Han, and R. M Hochstrasser. "Direct Observations of Ligand Dynamics in Hemoglobin by Subpicosecond Infrared-Spectroscopy". In: *Proc. Natl. Acad. Sci. USA* 86 (1989), pp. 8387–8391.

[170] M. Lim, T. A. Jackson, and P. A Anfinrud. "Mid-Infrared Vibrational Spectrum of CO after Photodissociation from Haem: Evidence for a Ligand Docking Site in the Haem Pocket of Haemoglobin and Myoglobin". In: *J. Chem. Phys.* 102 (1995), pp. 4355–4366.

[171] M. C. Thielges and M. D Fayer. "Protein Dynamics Studied with Ultrafast Two-Dimensional Infrared Vibrational Echo Spectroscopy". In: *Acc. Chem. Res.* 45 (2012), pp. 1866–1874.

[172] A. M. Woys et al. "A Strongly Absorbing Class of Non-Natural Labels for Probing Protein Electrostatics and Solvation with FTIR and 2D IR Spectroscopies". In: *J. Phys. Chem. B* 117 (2013), pp. 5009–5018.

[173] I. Peran et al. "General Strategy for the Bioorthogonal Incorporation of Strongly Absorbing, Solvation-Sensitive Infrared Probes into Proteins". In: *J. Phys. Chem. B* 118 (2014), pp. 7946–7953.

[174] J. T. King et al. "Crowding induced collective hydration of biological macromolecules over extended distance". In: *J. Am. Chem. Soc* 136 (2014), pp. 188–194.

[175] K.-I. Oh et al. "β-Azidoalanine as an IR Probe: Application to Amyloid Aβ(16-22) Aggregation". In: *J. Phys. Chem. B* 112 (2008), pp. 10352–10357.

[176] H. Taskent-Sezgin et al. "Azidohomoalanine: A conformationally sensitive IR probe of protein folding protein structure and electrostatics". In: *Angew. Chemie - Int. Ed.* 49 (2010), pp. 7473–7475.

[177] M. C. Thielges et al. "Two-Dimensional IR Spectroscopy of Protein Dynamics Using Two Vibrational Labels: A Site-Specific Genetically Encoded Unnatural Amino Acid and an Active Site Ligand". In: *J. Phys. Chem. B* 115 (2011), pp. 11294–11304.

[178] M. P. Wolfshorndl et al. "Covalently boundazido groups are very specific water sensors, even in hydrogen-bonding environments". In: *J. Phys. Chem. B* 116 (2012), pp. 1172–1179.

[179] S. Bagchi, S. G. Boxer, and M. D Fayer. "Ribonuclease S Dynamics Measured Using a Nitrile Label with 2D IR Vibrational Echo Spectroscopy". In: *J. Phys. Chem. B* 116 (2012), pp. 4034–4042.

[180] J. Zimmermann et al. "Cyano Groups as Probes of Protein Microenvironments and Dynamics". In: *Angew. Chem. Int. Ed.* 50 (2011), pp. 8333–8337.

[181] L. J. G. W. van Wilderen et al. "Vibrational dynamics and solvatochromism of the label SCN in various solvents and hemoglobin by time dependent IR and 2D-IR spectroscopy". In: *Phys. Chem. Chem. Phys.* 16 (2014), pp. 19643–19653.

[182] K. L. Kiick et al. "Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation". In: *Proc. Natl. Acad. Sci. USA* 99 (2002), p. 19.

[183] M. Simon, U. Zangemeister-Wittke, and A Plückthun. "Facile double-functionalization of designed ankyrin repeat proteins using click and thiol chemistries." In: *Bioconjugate Chem.* 23 (2012), pp. 279–286.

[184] B. Buchli et al. "Kinetic response of a photoperturbed allosteric protein." In: *Proc. Natl. Acad. Sci. USA* 110 (2013), pp. 11725–11730.

[185] S. A. Waldauer et al. "Effect of viscogens on the kinetic response of a photoperturbed allosteric protein". In: *J. Chem. Phys.* 141 (2014), p. 22D514.

[186] S. Buchenberg et al. "Long-rang conformational transition of a photoswitchable allosteric protein: A molecular dynamics simulation study". In: *J. Phys. Chem. B* 118 (2014), pp. 13468–13476.

[187] G. Kozlov, K. Gehring, and I Ekiel. "Solution structure of the PDZ2 domain from human phosphatase hPTP1E and its interactions with C-terminal peptides from the Fas receptors". In: *Biochemistry* 39 (2000), pp. 2572–2580.

[188] J. Zhang et al. "Crystallographic and Nuclear Magnetic Resonance Evaluation of the Impact of Peptide Binding to the Second PDZ Domain of Protein Tyrosine Phosphatase 1E". In: *Biochemistry* 49 (2010), pp. 9280–9291.

[189] E. J. Fuentes, C. J. Der, and A. L Lee. "Ligand-dependent dynamics and intramolecular signalling in a PDZ domain". In: *J. Mol. Biol.* 335 (2004), pp. 1105–1115.

[190] S. Gianni et al. "Demonstration of long-range interactions in a PDZ domain by NMR, kinetics and protein engineering." In: *Structure* 14 (2006), pp. 1801–1809.

[191] P. De Los Rios et al. "Functional Dynamics of PDZ Binding Domains: a Normal-Mode Analysis." In: *Biophys. J.* 89 (2005), pp. 14–21.

[192] K. Sharp and J. J Skinner. "Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling". In: *Proteins* 65 (2006), pp. 347–361.

[193] Z. N. Gerek and S. B Ozkan. "Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning." In: *PLoS Comput. Biol.* 7 (2011), e1002154.

[194] E. Cilia, G. W. Vuister, and T Lenaerts. "Accurate prediction of the dynamical changes within the second PDZ domain of PTP1e." In: *PLoS Comput. Biol.* 8 (2012), e1002794.

[195] G. M. Suel et al. "Evolutionarily conserved networks of residues mediate allosteric communication in proteins". In: *Nat. Struct. Biol.* 10 (2003), pp. 59–69.

[196] C. N. Chi et al. "Reassessing a sparse energetic network within a single protein domain." In: *Proc. Natl. Acad. Sci. USA* 105 (2008), pp. 4679–4684.

[197] D. C. Burns, F. Zhang, and G. A Woolley. "Synthesis of 3,3′-bis(sulfonato)-4,4′-bis(chloroacetamido)azobenzene and cysteine cross-linking for photo-control of protein conformation and activity." In: *Nat. Protoc.* 2 (2007), pp. 251–258.

[198] P. Hamm, R. A. Kaindl, and J Stenger. "Noise suppression in femtosecond mid-infrared light sources". In: *Opt. Lett.* 25 (2000), pp. 1798–1800.

[199] V. Volkov, R. Schanz, and P Hamm. "Active Phase Stabilization in Fourier-Transform Two-Dimensional Infrared Spectroscopy". In: *Opt. Lett.* 30 (2005), pp. 2010–2012.

[200] C. T. Middleton, D. B. Strasfeld, and M. T Zanni. "Polarization shaping in the mid-IR and polarization-based balanced heterodyne detection with application to 2D IR spectroscopy". In: *Opt. Express* 17 (2009), pp. 14526–14533.

[201] R. Bloem et al. "Enhancing signal detection and completely eliminating scattering using quasi-phase-cycling in 2D IR experiments". In: *Opt. Express* 18 (2010), pp. 27067–27078.

[202] P. J. M. Johnson, K. L. Koziol, and P Hamm. "Intrinsic phasing of heterodyne-detected multidimensional infrared spectra". In: *Opt. Express* 25 (2017), pp. 2928–2938.

[203] S. Pronk et al. "GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit". In: *Bioinformatics* 29 (2013), pp. 845–854.

[204] I. S. Joung and T. E Cheatham. "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.

[205] J. Wang et al. "Automatic atom type and bond type perception in molecular mechanical calculations". In: *J. Mol. Graph. Model.* 25 (2006), pp. 247–260.

[206] B. Hess et al. "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced and Scalable Molecular Simulation". In: *J. Chem. Theory Comput.* 4 (2008), pp. 435–447.

[207] O Steinhauser. "Reaction field simulation of water". In: *Mol. Phys.* 45 (1982), pp. 335–348.

[208] S. Wolf et al. "Infrared spectral marker bands characterizing a transient water wire inside a hydrophobic membrane protein". In: *J. Chem. Phys.* 141 (2014), p. 22D524.

[209] C. Lee, W. Yang, and R. G Parr. "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density". In: *Phys. Rev. B* 37 (1988), pp. 785–789.

[210] A. D Becke. "Density-Functional Thermochemistry III: the Role of Exact Exchange". In: *J. Chem. Phys.* 98 (1993), pp. 5648–5652.

[211] R. D Johnson III. "NIST 101. Computational Chemistry Comparison and Benchmark Database". In: *CCCBDBD Computational Chemistry Comparison and Benchmark Database* (1999.).

[212] R. S Mulliken. "Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I". In: *J. Chem. Phys.* 23 (1955), pp. 1833–1840.

[213] J. E. Moses and A. D Moorhouse. "The growing applications of click chemistry". In: *Chem. Soc. Rev.* 36 (2007), pp. 1249–1262.

[214] X. Pang and H.-X Zhou. "Rate constants and mechanisms of protein-ligand binding". In: *Annu. Rev. Biophys.* 46 (2017), pp. 105–130.

[215] U. Stelzl et al. "A human protein-protein interaction network: A resource for annotating the proteome". In: *Cell* 122 (2005), pp. 957–968.

[216] J. A McCammon. "Theory of biomolecular recognition". In: *Curr. Opin. Struct. Biol.* 8 (1998), pp. 245–249.

[217] M. R. Arkin, Y. Tang, and J. A Wells. "Small-molecule inhibitors of protein-protein interactions: profressing towards the reality". In: *Chem Biol* 21 (2014), pp. 1102–1114.

[218] R. E. Babine and S. L Bender. "Molecular recognition of protein-ligand complexes: Applications to drug design". In: *Chem. Rev.* 97 (1997), pp. 1359–1472.

[219] P. L. Kastritis and A. M. J. J Bonvin. "On the binding affinity of macromolecular interactions: daring to ask why proteins interact". In: *J. R. Soc. Interface* 10 (2013), p. 20120835.

[220] L. Zhao and J Chmielewski. "Inhibiting protein-protein interactions using designed molecules". In: *Curr. Opin. Struct. Biol.* 15 (2005), pp. 31–34.

[221] J Piehler. "New methodologies for measuring protein interactions in vivo and in vitro". In: *Curr. Opin. Struct. Biol.* 15 (2005), pp. 4–14.

[222] Daviter T. Eds Williams M. A. "Protein-Ligand Interactions Methods and Applications". In: *Springer: New York Heidelberg Dordrecht London, 2013.* ().

[223] N. J Greenfield. "Applications of circular dichroism in protein and peptide analysis". In: *TrAC - Trends Anal. Chem.* 18 (1999), pp. 236–244.

[224] T. Berggård, S. Linse, and P James. "Methods for the detection and analysis of protein-protein interactions". In: *Proteomics* 7 (2007), pp. 2833–2842.

[225] Z. Yan and J Wang. "Specificity quantification of biomolecular recognition and its implication for drug discovery". In: *Sci. Rep.* 2 (2012), pp. 1–7.

[226] R. E. Horness, E. J. Basom, and M. C Thielges. "Site-selective characterization of Src homology 3 domain molecular recognition with cyanophenylalanine infrared probes". In: *Anal. Methods* 7 (2015), pp. 7234–7241.

[227] R. E. Horness et al. "Resolution of site-specific conformational heterogeneity in proline-rich molecular recognition by Src homology 3 gomains". In: *J. Am. Chem. Soc.* 138 (2016), pp. 1130–1133.

[228] E. J. Basom, B. A. Manifold, and M. C Thielges. "Conformational heterogeneity mand the affinity of substrate molecular recognition by cytochrome P450cam". In: *Biochemistry* 56 (2017), pp. 3248–3256.

[229] B. Błasiak et al. "Vibrational probes: From small molecule solvatochromism theory and experiments to applications in complex systems". In: *Acc. Chem. Res.* 50 (2017), pp. 968–976.

[230] K.-I. Oh et al. "Beta-azidoalanine as an IR probe: application fo amyloid Abeta(16-22) aggregation". In: *J. Phys. Chem.* 112 (2008), pp. 10352–10357.

[231] P. Hamm and M. T Zanni. *Concepts and Methods of 2D Infrared Spectroscopy*. Cambridge University Press: Cambridge, 2011.

[232] A. Jakalian et al. "Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method". In: *J. Comput. Chem.* 21 (2000), pp. 132–146.

[233] A. Jakalian, D. B. Jack, and C. I Bayly. "Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation". In: *J. Comput. Chem.* 23 (2002), pp. 1623–1641.

[234] M. Ernst, S. Wolf, and G Stock. "Identification and validation of reaction coordinates describing protein functional motion: Hierarchical dynamics of T4 lysozyme". In: *J. Chem. Theory Comput.* (2017).

[235] S. W Hell. "Far-field optical nanoscopy". In: *Science* 316 (2007), pp. 1153–1158.

[236] Alan W Sousa da Silva and Wim F Vranken. "ACPYPE - AnteChamber PYthon Parser interfacE". In: *BMC research notes* 5 (2012), p. 367. ISSN: 1756-0500. DOI: 10.1186/1756-0500-5-367.

[237] Schrodinger LLC. The PyMOL Molecular Graphics System Version 1.8. In: *Schrödinger, LLC* (2015).

[238] 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.

[239] J Standfuss. "Membrane protein dynamics studied by X-ray lasers – or why only time will tell". In: *Curr. Opin. Struct. Biol.* 57 (2019), pp. 63–71.

[240] P. Hamm, J. Helbing, and J Bredenbeck. "Two-Dimensional Infrared Spectroscopy of Photoswitchable Peptides". In: *Annu. Rev. Phys. Chem.* 59 (2008), pp. 291–317.

[241] J. Bredenbeck et al. "Picosecond conformational transition and equilibration of a cyclic peptide". In: *Proc. Natl. Acad. Sci. USA* 100 (2003), pp. 6452–6457.

[242] H. Satzger et al. "Picosecond dynamics in water-soluble azobenzene-peptides." In: *Chem. Phys. Lett.* 396 (2004), pp. 191–197.

[243] A. Aemissegger et al. "A photoinducible beta-hairpin". In: *J. Am. Chem. Soc.* 127 (2005), pp. 2929–2936.

[244] S. Jurt et al. "A Photoswitchable Miniprotein Based on the Sequence of Avian Pancreatic Polypeptide". In: *Angew. Chem. Int. Ed.* 45 (2006), pp. 6297–6300.

[245] S. L. Dong et al. "A photocontrolled beta-hairpin peptide". In: *Chem. -A Eur. J.* 12 (2006), pp. 1114–1120.

[246] T. E. Schrader et al. "Light-triggered beta-hairpin folding and unfolding". In: *Proc. Natl. Acad. Sci. USA* 104 (2007), pp. 15729–15734.

[247] L. Lorenz et al. "Temperature- and Photocontrolled Unfolding/Folding of a Triple-Helical Azobenzene-Stapled Collagen Peptide Monitored by Infrared Spectroscopy". In: *ChemPhysChem* 17 (2016), pp. 1314–1320.

[248] F. Zhang et al. "Structure-based approach to the photocontrol of protein folding". In: *J. Am. Chem. Soc.* 131 (2009), pp. 2283–2289.

[249] B. Schierling et al. "Controlling the enzymatic activity of a restriction enzyme by light". In: *Proc. Natl. Acad. Sci. USA* 107 (2010), pp. 1361–1366.

[250] D Hoersch. "Engineering a light-controlled F 1 ATPase using structure-based protein design". In: *PeerJ* 4 (2016), e2286.

[251] B. Eisel et al. "Reversible optical control of F1Fo-ATP synthase using photo-switchable inhibitors". In: *FEBS Lett.* 592 (2018), pp. 343–355.

[252] R. S. Ritterson et al. "Design of a photoswitchable cadherin". In: *J. Am. Chem. Soc.* 135 (2013), pp. 12516–12519.

[253] A. A. Beharry et al. "Fluorescence imaging of azobenzene photoswitching in vivo". In: *Angew. Chemie - Int. Ed.* 50 (2011), pp. 1325–1327.

[254] J. Wachtveitl and A Zumbusch. "Azobenzene: An Optical Switch for in vivo Experiments". In: *ChemBioChem* 12 (2011), pp. 1169–1170.

[255] M. Borowiak et al. “Photoswitchable Inhibitors of Microtubule Dynamics Optically Control Mitosis and Cell Death”. In: *Cell* 162 (2015), pp. 402–411.

[256] D. Liu et al. “Site-specific incorporation of photoisomerizable azobenzene groups into ribonuclease S”. In: *Bioorganic Med. Chem. Lett.* 7 (1997), pp. 2677–2680.

[257] I. Hamachi et al. “Photoswitching of the Enzymatic Activity of Semisynthetic Ribonuclease S Bearing Phenylazophenylalanine at a Specific Site”. In: *Chem. Lett.* (1998), pp. 537–538.

[258] D. James, D. C. Burns, and G Woolley. “Kinetic characterization of ribonuclease S mutants containing photoisomerizable phenylazophenylalanine residues”. In: *Protein Eng. Des. Sel.* 14 (2001), pp. 983–991.

[259] S. Kneissl et al. “Photocontrollable Peptide-Based Switches Target the Anti-Apoptotic Protein Bcl-x L”. In: *ChemBioChem* 9 (2008), pp. 3046–3054.

[260] D. J. Boerema, V. A. Tereshko, and S. B. H. Kent. “Total synthesis by modern chemical ligation methods and high resolution (1.1 Å) x-ray structure of ribonuclease a”. In: *Peptide Science* 90 (2008), pp. 278–286.

[261] S. D. Stelea and T. A Keiderling. “Pretransitional Structural Changes in the Thermal Denaturation of Ribonuclease S and S Protein”. In: *Biophys. J.* 83 (2002), pp. 2259–2269.

[262] J. F. Dice et al. “Regulation of Catabolism of Microinjected Ribonuclease A”. In: *J. Biol. Chem.* 261 (1986), pp. 6853–6859.

[263] J. F Dice. “Microinjected ribonuclease A as a probe for lysosomal pathways of intracellular protein degradation”. In: *J. Protein Chem.* 7 (1988), pp. 115–127.

[264] A. R. Dias et al. “Enthalpies of formation of cis-azobenzene and trans-azobenzene”. In: *J. Chem. Thermodyn.* 24 (1992), pp. 439–447.

[265] M. Das et al. “Attempts to delineate the relative contributions of changes in hydrophobicity and packing to changes in stability of ribonuclease s mutants”. In: *Biochemistry* 44 (2005), pp. 5923–5930.

[266] R. B. Best, W. Zheng, and J. Mittal. “Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association.” In: *J. Chem. Theory Comput.* 10 (2014), pp. 5113–5124.

[267] J. L. F. Abascal et al. “A potential model for the study of ices and amorphous water: TIP4P/Ice”. In: *J. Chem. Phys.* 122 (2005), p. 234511.

[268] F. Sittel, T. Filk, and G. Stock. “Principal component analysis on a torus: Theory and application to protein dynamics”. In: *J. Chem. Phys.* 147 (2017), p. 244101.

[269] M. A Cooper. “Label-free screening of bio-molecular interactions”. In: *Anal. Bioanal. Chem.* 377 (2003), pp. 834–842.

[270] J. J. Rosa and F. M Richards. “Hydrogen Exchange from Identified Regions of the S-Protein Component of Ribonuclease as a Function of Temperature , pH and the Binding of S-Peptide”. In: *J. Mol. Biol.* 145 (1981), pp. 835–851.

[271] A. M Labhardt. “Kinetic circular dichroism shows that the S-peptide a-helix of ribonuclease S unfolds fast and refolds slowly”. In: *Proc Natl Acad. Sci, USA* 81 (1984), pp. 7674–7678.

[272] R. W Cowgill. "Fluorescence and the structure of proteins. III. Effects of denaturation on fluorescence of insulin and ribonuclease". In: *Arch. Biochem. Biophys.* 104 (1964), pp. 84–92.

[273] M. Noronha et al. "Tracking local conformational changes of ribonuclease A using picosecond time-resolved fluorescence of the six tyrosine residues". In: *Biophys. J.* 92 (2007), pp. 4401–4414.

[274] W. Kabsch and C. Sander. "Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features". In: *Biopolymers* 22 (1983), pp. 2577–2637.

[275] C. Baakman et al. "A series of pdb-related databanks for everyday needs". In: *Nucleic Acids Research* 43 (2014), pp. D364–D368.

[276] G. Schreiber and A. E Keating. "Protein binding specificity versus promiscuity". In: *Curr. Opin. Struct. Biol.* 21 (2011), pp. 50–61.

[277] T. C. B. McLeish, T. L. Rodgers, and M. R. Wilson. "Allostery without conformational change: modelling protein dynamics at multiple scales". In: *Phys. Biol.* 10 (2013), p. 056004.

[278] Ruth Nussinov and C.-J. Tsai. "Allostery without a conformational change? Revisiting the paradigm". In: *Curr. Opin. Struc. Biol.* 30 (2015), pp. 17–24.

[279] D. Thirumalai et al. "Symmetry, Rigidity and Allosteric Signaling: From Monomeric Proteins to Molecular Machines". In: *Chem. Rev.* 119.12 (2019), pp. 6788–6821.

[280] Arthur G. Palmer. "NMR characterization of the dynamics of biomacromolecules". In: *Chem. Rev.* 104.8 (2004), pp. 3623–3640.

[281] A Mittermaier and L E Kay. "New tools provide new insights in NMR studies of protein dynamics". In: *Science* 312.5771 (2006), pp. 224–228.

[282] Dominique Bourgeois and Antoine Royant. "Advances in kinetic protein crystallography". In: *Curr. Opin. Struct. Biol.* 15.5 (2005), pp. 538–547.

[283] Xavier Deupi and Brian K Kobilka. "Dynamics and Function". In: *Physiology* 25.5 (2010), pp. 293–303.

[284] Sven Brüschweiler et al. "Direct observation of the dynamic process underlying allosteric signed transmission". In: *J. Am. Chem. Soc.* 131.8 (2009), pp. 3063–3068.

[285] C. Hyeon and D. Thirumalai. "Mechanical unfolding of RNA hairpins". In: *PNAS* 102 (2005), pp. 6789–6794.

[286] F. Pontiggia et al. "Free energy landscape of activation in a signalling protein at atomic resolution". In: *Nat. Commun.* 6 (2015), p. 7284.

[287] Colin A. Smith et al. "Allosteric switch regulates protein-protein binding through collective motion". In: *Proc. Natl. Acad. Sci. U. S. A.* 113.12 (2016), pp. 3269–3274.

[288] Jinho Kim et al. "Rewiring of PDZ domain-ligand interaction network contributed to eukaryotic evolution." In: *PLoS Genetics* 8 (2012), e1002510.

[289] M Sheng and C Sala. "Pdz D Omains and the O Rganization of S Upramolecular C Omplexes". In: *Annu. Rev. Neurosci.* 24.1 (2001), pp. 1–29.

[290] Anthony B. Law, Ernesto J. Fuentes, and Andrew L. Lee. "Conservation of side-chain dynamics within a protein family". In: *J. Am. Chem. Soc.* 131.18 (2009), pp. 6322–6323.

[291] Kimberly A. Reynolds, Richard N. McLaughlin, and Rama Ranganathan. "Hot spots for allosteric regulation on protein surfaces". In: *Cell* 147.7 (2011), pp. 1564–1575. URL: http://dx.doi.org/10.1016/j.cell.2011.10.049.

[292] Tracy T. Cao et al. "A kinase-regulated PDZ-domain interaction controls endocytic sorting of the β2-adrenergic receptor". In: *Nature* 401.6750 (1999), pp. 286–290.

[293] Angelo Toto et al. "Understanding the role of phosphorylation in the binding mechanism of a PDZ domain". In: *Protein Eng. Des. Sel.* 30.1 (2017), pp. 1–5.

[294] Celestine N. Chi et al. "Ligand binding by PDZ domains". In: *BioFactors* 38.5 (2012), pp. 338–348.

[295] Brankica Jankovic et al. "Photocontrolling Protein–Peptide Interactions: From Minimal Perturbation to Complete Unbinding". In: *J. Am. Chem. Soc.* 141.27 (2019), pp. 10702–10710.

[296] Jens Bredenbeck, Jan Helbing, and Peter Hamm. "Continuous scanning from picoseconds to microseconds in time resolved linear and nonlinear spectroscopy". In: *Rev. Sci. Instrum.* 75.11 (2004), pp. 4462–4466.

[297] Yuan Feng, Ilya Vinogradov, and Nien-Hui Ge. "General noise suppression scheme with reference detection in heterodyne nonlinear spectroscopy". In: *Opt. Express* 25.21 (2017), p. 26262.

[298] A Barth and C Zscherp. "What vibrations tell us about proteins". In: *Q. Rev. Biophys.* 35 (2002), pp. 369–430.

[299] V Borisenko and G A Woolley. "Reversibility of conformational switching in light-sensitive peptides". In: *J. Photochem. Photobiol. A Chem.* 173 (2005), pp. 21–28.

[300] J Sabelko, J Ervin, and M Gruebele. "Observations of strange kinetics in protein folding". In: *Proc. Natl. Acad. Sci. USA* 96 (1999), pp. 6031–6036.

[301] M. P. Hobson and A. N. Lasenby. "The entropic prior for distributions with positive and negative values". In: *Mon. Not. R. Astron. Soc.* 298.3 (1998), pp. 905–908.

[302] Anand T. N. Kumar et al. "On the Rate Distribution Analysis of Kinetic Data Using the Maximum Entropy Method: Applications to Myoglobin Relaxation on the Nanosecond and Femtosecond Timescales". In: *J. Phys. Chem. B* 105.32 (2001), pp. 7847–7856.

[303] Víctor A. Lórenz-Fonfría and Hideki Kandori. "Transformation of time-resolved spectra to lifetime-resolved spectra by maximum entropy inversion of the laplace transform". In: *Appl. Spectrosc.* 60.4 (2006), pp. 407–417.

[304] Tobias Baumann et al. "Site-Resolved Observation of Vibrational Energy Transfer Using a Genetically Encoded Ultrafast Heater". In: *Angew. Chemie - Int. Ed.* 58.9 (2019), pp. 2899–2903.

[305] Simon Brandt et al. "Machine Learning of Biomolecular Reaction Coordinates". In: *J. Phys. Chem. Lett.* 9.9 (2018), pp. 2144–2150.

[306] Ushnish Sengupta and Birgit Strodel. "Markov models for the elucidation of allosteric regulation". In: *Phil. Trans. R. Soc. B* 373.1749 (2018), p. 20170178. ISSN: 0962-8436. DOI: 10.1098/rstb.2017.0178.

[307] Steffen Wolf et al. "Estimation of Protein-Ligand Unbinding Kinetics Using Non-Equilibrium Targeted Molecular Dynamics Simulations". In: *J. Chem. Inf. Model.* 59.12 (2019), pp. 5135–5147.

[308] Jungwoo Joh and Jesús A. Del Alamo. "A current-transient methodology for trap analysis for GaN high electron mobility transistors". In: *IEEE Trans. Electron Devices* 58.1 (2011), pp. 132–140.

[309] Ulrich Essmann et al. "A smooth particle mesh Ewald method". In: *J. Chem. Phys.* 103.19 (1995), pp. 8577–8593.

[310] Tianqi Chen and Carlos Guestrin. "XGBoost: A scalable tree boosting system". In: *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 13-17-Augu (2016), pp. 785–794.

[311] Abhinav Jain and Gerhard Stock. "Hierarchical folding free energy landscape of HP35 revealed by most probable path clustering". In: *J. Phys. Chem. B* 118.28 (2014), pp. 7750–7760.

[312] Ho-Jin Lee and Jie J Zheng. "PDZ domains and their binding partners: structure, specificity and modification". In: *Cell Commun. Signal.* 8 (2010), p. 8.

[313] P. Hamm et al. In: *Preparation* (2020).