Master equation model to predict energy transport pathways in proteins

Luis Valiño Borau,1 Adnan Gulzar,1 and Gerhard Stock1, a)

Biomolecular Dynamics, Institute of Physics, Albert Ludwigs University, 79104 Freiburg, Germany

(Dated: January 2, 2020)

Recent time-resolved experiment and accompanying molecular dynamics simulations allow to monitor the flow of vibrational energy in biomolecules. 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.

I. INTRODUCTION

Progress in time-resolved vibrational spectroscopy has made it possible to monitor the flow of biomolecular energy in space and time.1–5 Usually the energy is injected into the molecule via impulsive photoexcitation of some chromophore, e.g., a natural heme group1 or a molecular photoswitch such as azobenzene.5 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 vibrations3 or the azide group of unnatural amino acids can be employed.6 In this way, recent experiments have been able to observe the anisotropic flow of vibrational energy in heme proteins7 and PDZ3 domain.8 Protein energy transfer is believed to indicate the long-range propagation of conformational change, which provides the basis for allosteric communication.9–13

Accompanying experimental studies, biomolecular energy flow has been described by atomistic molecular dynamics (MD) simulations.9–19 Moreover, various network models of energy transport have been proposed,20–27 which typically aim to predict the energy flow between specific parts (usually residues) of a protein. In particular, Buchenberg et al.25,26 suggested a master equation

\[ \frac{dE_i(t)}{dt} = \sum_j [k_{ij}E_i(t) - k_{ji}E_j(t)], \]  

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,20 which is typically found for solvated biomolecules.3 The energy transport rates of the master equation can be obtained, e.g., from normal mode theory,25 equilibrium MD simulations,9 or nonequilibrium MD simulations.17 If available, also experimental data or quantum energy transfer calculations28,29 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.26 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,

\[ k_{ij} = \frac{D_B}{\langle x_{ij}^2 \rangle} \sqrt{\frac{f_j}{f_i}}, \]

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.30 The factor \( \sqrt{f_j/f_i} \) assures that the rates obey the detailed balance relation

\[ k_{ij}f_i = k_{ji}f_j. \]
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. 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. 26 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. 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. 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. 1.

Secondly, we want to exploit a key virtue of master equation (1), 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. Here, we perform extensive Monte Carlo Markov chain simulations to explore the energy transport pathways in TrpZip2 and PDZ3.

II. THEORY AND METHODS

A. MD simulations

Gulzar et al. recently presented extensive nonequilibrium MD simulations of the energy flow in proteins TrpZip2 (PDB entry 1LE1) and PDZ3 (PDB entry 1BE9). All MD simulations were performed using GROMACS package v2016.3 (Ref. 36), Amber99sb*ILDN forcefield, and TIP3P water. Following suitable equilibrium runs at \( T_0 = 300 \text{ K} \) (100 ns length for TrpZip2, 5×1 \( \mu \text{s} \) 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 (\( \sim 1 \text{ ps} \)) internal conversion, 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 \( \approx 2 \text{ eV} \), resulting in \( \Delta T \approx 600 \text{ K} \). 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 \( \delta t = 0.7 \text{ fs} \)) with only the solvent coupled to the thermostat (coupling constant \( \tau_T = 10 \text{ ps} \) ) represents an efficient and accurate strategy.

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 \( i \)th residue, \( E_{i}^{\text{kin}}(t) = \sum_j E_{i,j}^{\text{kin}}(t) \), 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 \( N_{\text{traj}} \) nonequilibrium trajectories,

\[
E_{i}(t) = \frac{1}{f_{i}N_{\text{traj}}} \sum_{n=1}^{N_{\text{traj}}} E_{i}^{\text{kin}}(n,t) - E_{i}^{\text{eq}},
\]

where \( f_{i} \) denotes the degrees of freedom of residue \( i \). Since \( E_{i}^{\text{eq}} = k_B T/2 \) 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

\[
E_T(t) = \frac{1}{M} \sum_{i=1}^{M} E_i(t).
\] (5)

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 \( \alpha \) helical protein HP36 (PDB ID 1UNC) was simulated at \( T_0 = 300 \text{ K} \), in order to compare to previous results at low temperatures (10 K). \(^{26}\) “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 \( \beta \)-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 \( \beta \)-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 \( \beta \)-sheet. In all cases, \( N_{\text{tra}} = 5 \) 000 nonequilibrium energy transport simulations were performed. As an overview, Table I comprises all simulated protein systems together with the type of heater residue.

Table I. List of considered molecular systems, type of heater residues, and resulting master equation parameters including inverse heating rate \( 1/k_h \), solvent cooling time \( 1/k_{ps} \), solvent back-transfer time \( 1/k_{bp} \), and contact times \( \tau_{i,j} = 1/k_{i,j} \) (all in units of ps).

<table>
<thead>
<tr>
<th>System</th>
<th>Ala12</th>
<th>AlaZip</th>
<th>TrpZip2</th>
<th>HP36</th>
<th>PDZ3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heater</td>
<td>Val3</td>
<td>Val3</td>
<td>Azu3</td>
<td>Len16</td>
<td>Azu(-5)</td>
</tr>
<tr>
<td>( 1/k_h )</td>
<td>1.7</td>
<td>1.7</td>
<td>5.9</td>
<td>3.1</td>
<td>5.9</td>
</tr>
<tr>
<td>( 1/k_{ps} )</td>
<td>7.9</td>
<td>8.7</td>
<td>8.3</td>
<td>8.8</td>
<td>10</td>
</tr>
<tr>
<td>( 1/k_{bp} )</td>
<td>190</td>
<td>250</td>
<td>120</td>
<td>120</td>
<td>60</td>
</tr>
<tr>
<td>contact times</td>
<td>( \tau_{0,10} = 21 )</td>
<td>( \tau_{0,10} = 125 )</td>
<td>( \tau_{4,15} = 9.1 )</td>
<td>( \tau_{7,32} = 50 )</td>
<td>( \tau_{7,32} = 50 )</td>
</tr>
<tr>
<td>( \tau_{7,10} = 10 )</td>
<td>( \tau_{7,10} = 5.9 )</td>
<td>( \tau_{18,26} = 17 )</td>
<td>( \tau_{7,10} = 10 )</td>
<td>( \tau_{7,32} = 50 )</td>
<td></td>
</tr>
<tr>
<td>( \tau_{7,12} = 31 )</td>
<td>( \tau_{7,12} = 43 )</td>
<td>( \tau_{7,12} = 17 )</td>
<td>( \tau_{7,12} = 7.7 )</td>
<td>( \tau_{7,12} = 17 )</td>
<td></td>
</tr>
<tr>
<td>( \tau_{7,8} = 13 )</td>
<td>( \tau_{7,8} = 9.1 )</td>
<td>( \tau_{7,8} = 9.1 )</td>
<td>( \tau_{7,8} = 9.1 )</td>
<td>( \tau_{7,8} = 9.1 )</td>
<td></td>
</tr>
<tr>
<td>( \tau_{7,5} = 71 )</td>
<td>( \tau_{7,5} = 71 )</td>
<td>( \tau_{7,5} = 71 )</td>
<td>( \tau_{7,5} = 71 )</td>
<td>( \tau_{7,5} = 71 )</td>
<td></td>
</tr>
</tbody>
</table>

B. 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 \(^{45}\) 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 (1), all these processes are modeled by transport rates \( k_{ij} \equiv k_{i\rightarrow j} \). 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 \( k_{0i} = k_h \delta_{in} \) and corresponding back rates \( k_{pi} \) according to Eq. (3).

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 \( k_{i,N+1} = k_{ps} \) for \( i = 0, \ldots, N \). To reach thermal equilibrium at long times, we also need to invoke a small back-rate \( k_{N+1,i} = k_{bp} \). While more sophisticated models can be considered, \(^{44} \) (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), 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 \( \langle x_i^2 \rangle \) is readily obtained from the molecular structure. In principle, \( D_B \) may depend on the secondary structure (e.g., \( \alpha \)-helical, turn or \( \beta \)-sheet) and on temperature.

C. 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. \(^{26} \) proposed a harmonic model
that yields for the contact transport rate

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

(6)

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. 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, 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 intrastand hydrogen bonds between $\beta$-sheets lead to relatively modest contact transport rates. To allow for a single transport constant $B_C$, Buchenberg et al. 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) for backbone transport, we then obtain for the contact transport rate

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

(7)

where $D_C$ denotes the contact diffusion constant and $\langle \delta q_{ij}^2 \rangle$ represents the mean square distance of the two contact atoms [rather than the corresponding variance as in Eq. (6)]. In fact, Reid et al. has recently found for myoglobin that the rates of various charged contacts rather obey Eq. (7) than Eq. (6). 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 geometricaly 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 $\beta$-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 $P$.

D. Parameterization

As discussed above, the parameterization of master equation (1) 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 $E_i(t)$ of the master equation and the corresponding energies $E_i^{\text{MD}}(t)$ from the nonequilibrium MD data,

$$\text{RMSD} = \left[ \sum_{i, t} \left( \frac{E_i^{\text{MD}}(t) - E_i(t)}{E_i^W(t)} \right)^2 \right]^{1/2}.$$  (8)

Here the sum runs over all residues $i = 0, \ldots, N+1$ and over all simulation times $t = t_{\text{min}} + n\delta t$, where we choose $\delta t = 0.0175$ ps to include $10^4$ data points in the fit for each residue, and $t_{\text{min}} = 0.1$ ps to discard extreme energy fluctuations of the initial phase of the nonequilibrium MD simulations. Proceeding this way, Buchenberg et al. 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 $\alpha$ and $\beta$ 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. (3) 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. (5). 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. II B) and the backbone diffusion coefficient \( D_B \) (defined in Eq. (2)), 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., \( \leq 4.5 \) Å), we obtain a contact map of the system.\(^{45}\) 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\(^{26}\) 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. (6) and (7). 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. 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 \( k_h \), 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 \( (8), \) 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.

E. 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 (1) 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 \( \rightarrow 2 \rightarrow 3 \) and 1 \( \rightarrow 2 \rightarrow 4 \rightarrow 2 \rightarrow 3 \) would be combined into a single pathway 1 \( \rightarrow 2 \rightarrow 3 \) 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.
### III. RESULTS AND DISCUSSION

To derive a general and consistent set of parameters that determine the rate matrix \( \{ k_{ji} \} \) of master equation (1), 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 \( \alpha \)-helical protein HP36. Next we consider two types of \( \beta \)-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.

#### A. 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 shows the time evolution of the residue energies \( E_j(t) \) [Eq. (4)] 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.

![Figure 2](image.png)

Figure 2. Time evolution of residue energies [Eq. (4)] 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).

In master equation (1), 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 \( k_{sp} \) and \( k_{sp} \). Adopting \( 1/k_h = 1.7 \text{ ps}, D_B = 1.1 \text{ nm}^2/\text{ps}, 1/k_{sp} = 7.9 \text{ ps} \) and \( 1/k_{sp} = 210 \text{ ps} \), Fig. 2 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., \( 1/k_h = 2.5 \text{ ps} \) and \( D_B = 1.4 \text{ nm}^2/\text{ps} \), see Fig. S2). Here we choose \( D_B = 1.1 \text{ nm}^2/\text{ps} \) which best approximates the observed peak time of the residue energies, and \( 1/k_h = 1.7 \text{ ps} \) as corresponding best fit. Insertion of \( D_B = 1.1 \text{ nm}^2/\text{ps} \) in Eq. (2) yields an energy transport time between two adjacent alanine residues of \( \approx 0.43 \text{ ps} \), which is in line with previous work.\(^{17}\)

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. II D. Using various values of \( k_{ps} \) and \( k_{sp} \), Fig. S3 reveals that we get reasonable results for \( 7.5 \text{ ps} \leq 1/k_{ps} \leq 8.5 \text{ ps} \) and \( 150 \text{ ps} \leq 1/k_{sp} \leq 300 \text{ ps} \), showing that particularly the value of the back rate is not very critical. Note the factor \( k_{ps}/k_{sp} \approx 30 \) reflects the different number of degrees of freedom of protein and solvent, respectively.

Figure 2 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 \( \alpha \)-helical protein villin headpiece (HP36) represents another molecular system whose energy flow is dominated by backbone transport. In contrast to previous work on HP36,\(^{26}\) 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. II A). Nonetheless, quite similar to the results of Ref. 26, 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,\textsuperscript{26} 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 $k_h = 1/3.1 \text{ ps}$, a backbone diffusion coefficient $D_B = 1.1 \text{ nm}^2/\text{ps}$, solvent dissipation rates $k_{ps} = 1/8.8 \text{ ps}$ and $k_{sp} = 1/120 \text{ ps}$, and contact rates $k_{4,15} = 1/9.2 \text{ ps}$ (a triple contact evolving from a strong salt bridge between Asp4 and Arg15) and $k_{18,26} = 1/14 \text{ ps}$ (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 I. In line with a previous study,\textsuperscript{3} $D_B = 1.1 \text{ nm}^2/\text{ps}$ appears to be the same for extended structures and for $\alpha$-helical structures, meaning that the hydrogen bonds of the $\alpha$-helix do not significantly contribute to the energy transfer. Moreover, this choice of $D_B$ is quite close to the value of $D_B = 1.25 \text{ nm}^2/\text{ps}$ found by Buchenberg et al.\textsuperscript{26} As the latter study was conducted at 10 K rather than 300 K, the backbone diffusion coefficient appears to depend only little on temperature.

B. Contact transport across a $\beta$-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 $\beta$-hairpin, see Fig. 1. Although the system may become unstable at longer times (see Sec. II A), it remains in a stable $\beta$-hairpin structure during the 50 ps long nonequilibrium MD runs. Figure 3 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. 2, moreover find significant enhancement of the residue energies at large sequence distances from the heater residue Val3 and residues Ala10 and Ala12 (Fig. 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). This yields for contact (1,12) the interstrand hydrogen bond contact rate $k_{1,12} = 1/17 \text{ ps}$ and the heater contact rate $k_{0,12} = 1/21 \text{ ps}$, as well as for contact (3,10) the rates $k_{3,10} = 1/10 \text{ ps}$ and $k_{0,10} = 1/31 \text{ ps}$. Since the rate of contact (5,8) is not well defined, we adopted a value of $k_{5,8} = 1/13 \text{ ps}$ which is in line with the squared distance scaling rule discussed below. Figure 3 reveals that the resulting model reproduces the MD result with similar quality than the simpler model neglecting heater contacts. As discussed in Sec. III C, 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 $\beta$-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.\textsuperscript{46} We

![Figure 3: Time evolution of residue energies of AlaZip.](image)
first focus on a TrpZip2 mutation with the Azu heater positioned in residue 3 (Fig. 1), which was recently studied using nonequilibrium MD simulations.\textsuperscript{31} 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. 4 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 $\beta$-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. (4)] of TrpZip2 are generally smaller, because the residues of TrpZip2 contain more degrees of freedom.

![Figure 4](image)

Figure 4. 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 $k_h = 1/5.9 \text{ ps}$, solvent dissipation rates $k_{ps} = 1/8.3 \text{ ps}$ and $k_{sp} = 1/120 \text{ ps}$, and a backbone diffusion coefficient $D_B = 1.1 \text{ nm}^2/\text{ps}$. The contact rates are $k_{a12} = 1/43 \text{ ps}$, $k_{112} = 1/7.7 \text{ ps}$, $k_{a9,10} = 1/130 \text{ ps}$ $k_{3,10} = 1/5.9 \text{ ps}$, $k_{5,8} = 1/9.1 \text{ ps}$ and $k_{9,5} = 1/71 \text{ ps}$. Figure 4 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.

C. 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 ($D_B = 1.1 \text{ nm}^2/\text{ps}$), a heating rate that reflects the size of the respective heater side-chain (e.g., $k_h = 1/5.9 \text{ ps}$ for Azu), and a typical cooling time of $1/k_{ps} = 8.9 \text{ ps}$, see Table I. 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.,\textsuperscript{31} 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 $\beta$-hairpins exhibit intrastand 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 (6) and (7), which relate the rates to the variance $\delta q_{ij}^2$ and the squared distance $q_{ij}$ of the corresponding contact, respectively.

![Figure 5](image)

Figure 5. Scaling rules of contact transport. Shown are contact rates for interstrand hydrogen bonds (1,12) and (3,10) of the five studied $\beta$-hairpins (AlaZip, and mutations M1-M4 of TrpZip2), as well as most relevant contact rates found for HP36 and PDZ3. (a) Following scaling rule (6), 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. (6) 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. (7).
Depicting these rates as a function of $\sqrt{f_j\langle\delta q_{ij}^2\rangle}$ and $\sqrt{f_j/f_i\langle\delta q_{ij}^2\rangle}$, Fig. 5 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 (6) 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 $\cdot 10^{-5}$ nm$^2$/ps. As discussed in Sec. II C and Appendix A, this reflects the fact that the rate (A4) 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.,$^{27}$ 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.$^{26}$ 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 (7) depicted in Fig. 5b, on the other hand, is seen to describe all available contact rates with a single contact diffusion constant, $D_C = 2.1 \cdot 10^{-3}$ nm$^2$/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.

D. 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. II E).

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 $\beta$-hairpin (Fig. 1). To this end, Fig. 6 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 ($1/k_{3,10} = 10$ ps) and a heater contact ($1/k_{0,10} = 21$ 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.

![Figure 6](image.png)

Figure 6. 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 ($\beta$C), heater contacts (HC), and other polar contacts (PC).

The situation is somewhat more complex for TrpZip2, which contains diverse residues with in part large side chains. We find strong interstrand hydrogen bonds ($1/k_{3,10} = 5.9$ ps) that represent the most effective pathway (27 %), followed by two heater contact pathways, either directly ($1/k_{0,10} = 125$ ps, 14 %) or via Lys12 ($1/k_{0,12} = 43$ ps, 23 %). Backbone transport contributes only 15 % of the energy. This is because several backbone transport steps are relatively slow, e.g., $\sim 1.5$ ps for Trp4$\rightarrow$Glu5, Lys8$\rightarrow$Trp9 and Trp9$\rightarrow$Aha10. This is a consequence of the large size of the involved residues and the associated long interatomic distances [cf. Eq. (2)]. Moreover, the succession of a large and a small residue...
causes a “bottleneck” for the energy flow, which is described in Eq. (2) by the detailed balance factor $\sqrt{f_i/f_j}$.

E. 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,\textsuperscript{19,22,24} for which first experimental results\textsuperscript{8} as well as detailed nonequilibrium MD data are available.\textsuperscript{31} 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\textsuperscript{47} indicated the population of a single conformation with the Azu residue turning to the $\beta_2$-sheet, here we restrict the analysis to this state.

Figure 7 shows the energy flow in PDZ3 as predicted by the MD simulations.\textsuperscript{34} Since the Azu heater is attached to the N-terminus of the ligand, the energy propagates through the ligand backbone and successively reaches all of its residues within a few picoseconds. Subsequently, the energy transfers via non-covalent contacts to the protein, where it reaches the $\alpha_2$-helix and the $\beta_2$-sheet, as well as a few more distant residues. On the $\alpha_2$ side, the ligand may form hydrogen bonds with His372, which is seen to receive some energy. On the $\beta_2$ 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 ($D_B = 1.1 \text{ nm}^2/\text{ps}$), the heating rate of Azu ($k_h = 1/5.9 \text{ ps}$), and determined solvent dissipation rates $k_{sp} = 1/10 \text{ ps}$ and $k_{ap} = 1/60 \text{ ps}$. 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 $\beta$-sheets of PDZ3, whose rates were obtained using scaling rule (7), see Table S3. Moreover, we identified 4 heater contacts with surrounding residues, 2 hydrogen bonds between ligand and $\beta_2$ sheet, a hydrogen bond between Gly329 and His372, as well as a nonpolar contact involving the beta-CH2 groups of residues Lys(-4) and His372, see Table I for the associated rates. Employing these rates to test the applicability of scaling rules (6) and (7), Fig. 5 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 7 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 7 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.

IV. CONCLUSIONS

We have outlined a general master equation model that describes the energy transport in proteins. It relies on scaling rules (2) and (7) 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\textsuperscript{8,48} and simulation,\textsuperscript{16,31} this cooling of the protein in water occurs on a timescale of 8 - 10 ps at 300 K (Table I). 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), which requires only an overall backbone diffusion coefficient ($D_B = 1.1 \text{ nm}^2/\text{ps}$) 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
Figure 7. Energy transfer in PDZ3. Top: Time evolution of residue energies, reflecting the energy flow from the ligand via α2 helix and β2-sheet to more remote regions. Bottom left: Close-up view of the binding pocket, including the ligand (green) and the adjacent α2 and β2 regions of the protein (red). By convention, the ligand is labeled from -5 to 0, while the protein is labeled 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).

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 I), which renders contact transport at least one order of magnitude slower than backbone transport.

4. The previously suggested inverse variance rule\textsuperscript{26} of contact transport was shown to depend on the way the contact atoms are connected to the rest of the protein (see Appendix A), which limits the practical use of the theory. The new inverse square distance scaling rule (7), on the other hand, was found to describe virtually all available contact rates with a single contact diffusion constant, $D_C = 2.1 \cdot 10^{-3} \text{nm}^2/\text{ps}$. 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.\textsuperscript{49}

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) and (7) 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.

Supplementary Material

Contains computational methods, conformational distribution of Ala12, dependence of energy transport on heater rate, backbone diffusion coefficient and cooling rate, energy transport in HP36 and three other mutants of TrpZip2, master equation parameters of these mutants, and contact rates of PDZ3.

Acknowledgment

We thank Jens Bredenbeck and his group for helpful discussions and for sharing experimental data prior to publication, and Sebastian Buchenberg and Steffen Wolf for numerous instructive discussions. This work has been supported by the Deutsche Forschungsgemeinschaft (Sto 247/10-2), the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen and the Rechenzentrum of the University of Freiburg, the state of Baden-Württemberg through bwHPC and the DFG through grants no INST 37/935-1 FUGG (RV bw16I016) and no INST 39/963-1 FUGG (RV bw18A004), and the Black Forest Grid Initiative.

Appendix A: Harmonic model of contact transport

To describe the through-space energy transport via a contact between two protein residues, Fig. 8 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

\[
\ddot{q}_1 = -d(q_1 - q_2) - D_L(q_1 - q_L), \quad (A1)
\]

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 \( q_1 \) 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 8. Scheme of the one-dimensional harmonic model of contact energy transport.](image)

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

\[
\ddot{q} + (D + 2d)q = 0, \quad \ddot{Q} + DQ = 0, \quad (A2)
\]

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. (A2), 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

\[
E_2(t) = \frac{\ddot{q}_2^2}{2} = \frac{v_0}{2} \sin^2 \left( \frac{\omega + \Omega}{2} \right) \sin^2 \left( \frac{\omega - \Omega}{2} \right). \quad (A3)
\]

Averaging over the rapidly oscillating term with \( \omega + \Omega \), we focus on the slowly varying term \( \sin^2 \left( \frac{\omega - \Omega}{2} \right) \), which reaches its maximum at \( (\Omega - \omega)t = \pi \). If we define the energy rise time \( \tau_{12} = 1/\tau_{12} = 2/\pi (\omega - \Omega) = 2/\pi (\sqrt{D + 2d} - \sqrt{D}) \). 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

\[
k_{12}^C = \frac{2}{\pi} \frac{d}{\sqrt{D}}. \quad (A4)
\]

Since for a harmonic oscillator \( \langle V \rangle = d/2 \langle \delta q^2 \rangle = \langle E_{\text{kin}} \rangle = k_B T/2 \), the spring constant \( d \) can be calculated from the variance of the displacement \( \delta q = (q - \langle q \rangle) \) via \( d = k_B T/\langle \delta q^2 \rangle \). Using \( B_C = 2k_B T/(\pi \sqrt{D}) \) then leads to scaling rule (6). Note that the model predicts that rigidly connected contacts (with a large spring constant \( D \)) lead to smaller rates than loosely bound contacts.

REFERENCES

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this typeset version once it has been copyedited and

PLEASE CITE THIS ARTICLE AS DOI: 23


Although the concept of a local temperature is difficult to define in a strict sense, for the sake of a simple physical picture we invoke the kinetic energy of an atom group as a measure of local temperature.


M. Ernst, F. Sittel, and G. Stock, Contact- and distance-based

46 J. Bredenbeck et al., in preparation.

