Cooperative protein allosteric transition mediated by a fluctuating transmission network

Matthias Post, Benjamin Lickert, Georg Diez, Steffen Wolf,* and Gerhard Stock†
Biomolecular Dynamics, Institute of Physics, Albert Ludwigs University, 79104 Freiburg, Germany.
(Dated: April 28, 2022)

Allosteric communication between distant protein sites represents a key mechanism of biomolecular regulation and signal transduction. Compared to other processes such as protein folding, however, the dynamical evolution of allosteric transitions is still not well understood. As example of allosteric coupling between distant protein regions, we consider the global open-closed motion of the two domains of T4 lysozyme, which is triggered by local motion in the hinge region. Combining extensive molecular dynamics simulations with a correlation analysis of interresidue contacts, we identify a network of interresidue distances that move in a concerted manner. The cooperative process originates from a cogwheel-like motion of the hydrophobic core in the hinge region, which constitutes a flexible transmission network. Through rigid contacts and the protein backbone, the small local changes of the hydrophobic core are passed on to the distant terminal domains and lead to the emergence of a rare global conformational transition. As in an Ising-type model, the cooperativity of the allosteric transition can be explained via the interaction of local fluctuations.

INTRODUCTION

Proteins may coexist in various conformational states of different function, e.g., in an active and an inactive state. Transitions between these states can occur spontaneously or may be triggered by an external stimulus such as ligand binding. A prime example is allosteric communication, where a binding event at one site of a protein may affect the structure and dynamics of another distant site.[1–10] Even without the presence of a ligand, many allosteric systems thermally populate both protein states, thus representing a bistable system. Common to all cases is that the associated structural rearrangement requires the interaction of distant regions of the system. Although these long-range or allosteric couplings appear to be ubiquitous in protein systems,[1] surprisingly little is known about the dynamical mechanism, its microscopical details, and the time evolution of the associated conformational transition.[8] We note that the situation is quite different for the protein folding problem, where several decades of theoretical and experimental work have resulted in a quite well-established picture how folding and unfolding proceed. This includes general scenarios such as cooperative two-state and multistate downhill folding,[11] as well as dynamical mechanisms such as zipping or diffusion limited processes.[12] The dynamical process underlying an allosteric transition, on the other hand, appears much more elusive due to the typically small local structural changes, which are quite challenging to observe in experiments.[13–15] While allosteric transitions have been rarely witnessed in unbiased molecular dynamics (MD) simulations,[9,16–20] numerous residue-based network models have been suggested, which aim to predict intramolecular signaling.[7,21] Furthermore Markov state models,[22] which describe the conformational dynamics via memory-less jumps between metastable states, have been applied to allostery.[17,23]

In this work we consider T4 lysozyme (T4L) as a well-studied example of a bistable two-domain protein.[24–29] Aiming to destroy bacterial cell walls by catalyzing the cleavage of glycosidic bonds, T4L performs an open-closed transition of its two domains that resembles a Pac-Man, see Fig. 1a. This motion of the "mouth" region was recently shown to be triggered by a locking mechanism in the "hinge" region of T4L, revealing that these two distant regions are allosterically coupled.[29] (Here we use the term 'allostery' for general long-range communication, independent from ligand binding.) With lifetimes of the open and closed state in the order of a few microseconds, T4L is one of the few protein systems, whose large-scale conformational transitions can be sufficiently sampled by unbiased simulations. Adopting the 61 µs-long all-atom MD simulation of Ernst et al.,[29] here we elucidate the microscopic mechanism of this long-range coupling. Previous work showed that the identification of the underlying reaction coordinates poses a challenge to various dimensionality reduction approaches.[28–31]

Combining a correlation analysis of contact features[32] with a careful inspection of individual transitions, we show that the open-closed transition of T4L amounts to a cooperative process, where a large number of interatomic distances from the hinge to the mouth region act in a simultaneous manner. In particular, we find that the hydrophobic core of the hinge region constitutes a flexible transmission network, which represents the molecular machinery at the heart of the open-closed transition.

RESULTS AND DISCUSSION

Two-dimensional model: the locking mechanism. The open-closed motion of T4L is usually described by the mouth opening width x (Fig. 1a), which can be represented, e.g., by the distance $d_{20,145}$ between residues 20 and 145 of the N- and C-domain, respectively. Employing the MD trajectory by Ernst et al.,[29] Fig. 1c reveals that the open-closed transitions of $x(t)$ occur on a microsecond timescale, while the transition

path time lasts only nanoseconds. From the data we estimate equilibrium populations of 70 % (open) and 30 % (closed) and mean waiting times of $\tau_{\mathrm{o}\to\mathrm{c}} \sim 4\,\mu\mathrm{s}$ and $\tau_{\mathrm{c}\to\mathrm{o}} \sim 2\,\mu\mathrm{s}$, which is in excellent agreement with recent experimental results.[27] The corresponding free energy profile $\Delta G(x)$, however, exhibits a rather low energy barrier ($\Delta G^{\neq} \sim 1 k_{\mathrm{B}}T$). Adopting the standard expression for the transition rate, $k = 1/\tau = k_0\, e^{-\Delta G^{\neq}/k_{\mathrm{B}}T}$, it is obvious that the rate is almost solely caused by the prefactor k_0.[33] This indicates that the opening width x does not reveal the underlying mechanism of the process and is therefore not a suitable one-dimensional reaction coordinate.

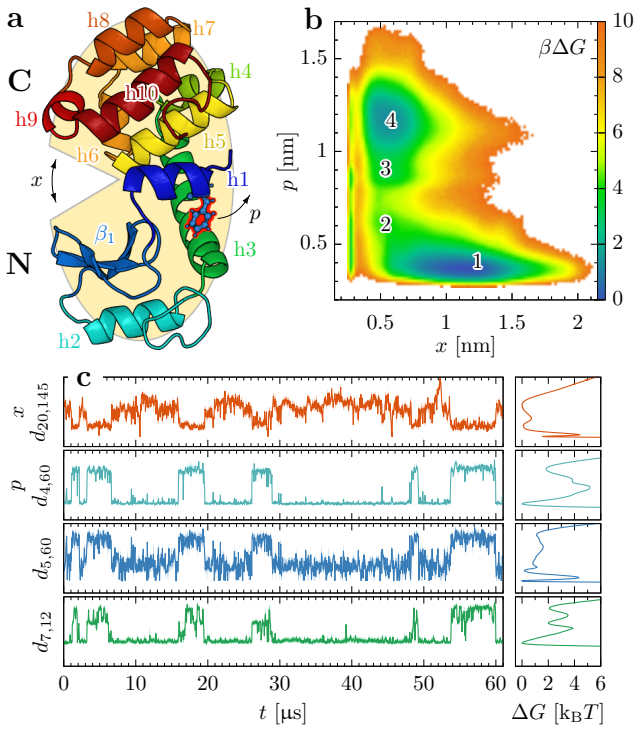

FIG. 1. (a) Structure of the N- and C-domain of T4L, indicating the opening coordinate x of the "mouth" region and the locking coordinate p that describes the motion of key residue Phe4 (red) in the "hinge" region. (b) Two-dimensional free energy landscape $\Delta G(x,p)$ (in units of $k_{\mathrm{B}}T$), indicating four conformational states. (c) MD time traces of various distance coordinates describing the open-closed transition and their corresponding free energy profiles.

Indeed, when we employed targeted MD simulations to enforce the open-close motion by pulling along x, the two-state behavior of T4L could not be recovered.[29] Rather, the transition was found to be triggered by a locking mechanism, by which the side chain of Phe4 changes from a solvent-exposed to a hydrophobically-buried state. Describing the motion of Phe4 via the hydrophobic contact distance between Phe4 and Lys60, $d_{4,60} \equiv p$, this coordinate discriminates between locked ($p \lesssim 0.7$, hydrophobically buried) and free ($p \gtrsim 0.7$, solvent exposed) states. Figures 1b,c show the free energy landscape $\Delta G(x,p)$ and the time traces $x(t)$, $p(t)$ of the resulting two-dimensional model of T4L. The landscape exhibits four metastable conformational states labeled by numbers **1** - **4**, including the prominent open locked state **1** and closed free state **4**, as well as two sparsely populated intermediate closed states **2** and **3**.

The locking mechanism is found to stabilize the open and closed states and can therefore be considered as the origin of the bistable behavior of T4L. Demonstrating the coupling of opening coordinate x and locking coordinate p, the free energy landscape $\Delta G(x,p)$ also provides first mechanistic insights of the conformational transition.[29] In particular $\Delta G(x,p)$ indicates that most transitions proceed along the route **1**→**2**→**3**→**4** and back along **4**→**3**→**2**→**1**. On the other hand, the two-dimensional model of T4L apparently still does not include some important aspects of the dynamics, because the corresponding highest free energy barrier is still relatively small ($\sim 5\,k_{\mathrm{B}}T$). As we show in this work, the missing link is the long-range propagation of conformational change, which couples the (un)locking of Phe4 in the hinge region to the opening/closing of the mouth region.

Correlation analysis: cooperative dynamics. To describe the mechanistic coupling between the mouth and hinge regions in T4L, we want to identify internal coordinates that change significantly during the open-closed transition. Following Ernst et al.,[29] we consider interresidue contact distances and side-chain χ_1 dihedral angles, as they report on near-order interactions, from which long-distance changes result as a consequence. Assuming that a contact is formed if the distance d_{ij} between the closest non-hydrogen atoms of residues i and j is shorter than 4.5 Å,[34] we identified 844 interresidue contacts that are formed at least 1 % of the simulation time. Moreover, we focused on the first side-chain dihedral angle of each residue, resulting in 131 coordinates χ_n (Ala, Gly and Pro do not exhibit or change this angle). To study which of these coordinates $\{x_\alpha\} = \{d_{ij}, \chi_n\}$ are interrelated to each other, we calculated the linear correlation matrix

$$\sigma_{\alpha\beta} = \frac{\langle \delta x_\alpha\, \delta x_\beta \rangle}{\langle \delta x_\alpha^2 \rangle^{1/2}\, \langle \delta x_\beta^2 \rangle^{1/2}}, \qquad (1)$$

where $\delta x_\alpha = x_\alpha - \langle x_\alpha \rangle$ and $\langle \ldots \rangle$ denotes the time average over the MD data. Aiming to discriminate collective motions underlying functional dynamics from uncorrelated motions, we wish to rearrange this matrix in an approximately block-diagonal form. Employing the Python package MoSAIC,[32] this is achieved via a community detection technique called Leiden clustering.[35]

The resulting block-diagonalized correlation matrix in Fig. 2a shows a main cluster (i.e., the first block) that contains 85 strongly correlated coordinates, which account for the open-closed transition (see below). Moreover, we obtain six smaller clusters describing local motions that are not related to the open-closed motion (Fig. S1), and the remaining 530 clusters that exhibit

only minor correlation and are therefore considered as noise (see Table S1 for a list of all clusters and coordinates). The latter finding is associated with the fact that in T4L intraprotein contacts are quite stable and only fluctuate around their mean distance, while contacts on the protein surface frequently form and break and hence fluctuate randomly. That is, of the 976 considered internal coordinates only 85 (82 contact distances and 3 dihedral angles) are in fact associated with the open-closed transition of T4L.

Displaying these 82 highly correlated distances in the molecular structure of T4L, Fig. 2b shows that the coordinates span the space connecting the mouth and hinge regions. They include 36 mouth-opening coordinates and 15 coordinates involving Phe4, while the remaining coordinates are mostly located in the hinge region. As demonstrated by the time evolution of some important examples (Fig. 1c), these coordinates follow the microsecond open-closed transitions of T4L. Hence, the above correlation analysis indicates that the open-closed transition is mediated via the 82 highly correlated contact distances that connect the hinge and mouth regions of T4L.

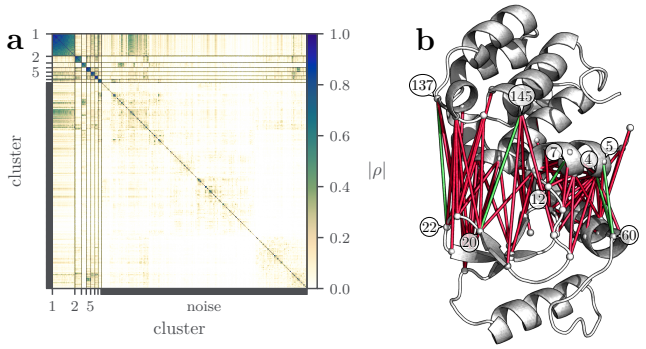

FIG. 2. (a) Block-diagonal form of the correlation matrix built from 844 contact distances and 131 side-chain dihedral angles of T4L. The first block (cluster 1) contains 85 highly correlated coordinates that account for the open-closed transition. (b) Illustration of the 82 relevant contact distances in red, as well as in green the important distances $d_{4,60}$, $d_{5,60}$, $d_{22,137}$, $d_{20,145}$ describing the opening of the mouth, and $d_{7,12}$ accounting for the (de)stabilization of helix 1.

While the overall time evolutions of the coordinates in Fig. 1c are clearly correlated, the various contact distances are seen to differ in their fluctuations in the open and closed state. E.g., opening coordinate x fluctuates considerably in the open state and only little in the closed state, while it is the other way round for the locking coordinate p. This is because amino acid side-chains that are in contact generally fluctuate significantly less than in the case of a broken contact. Although the fast fluctuations of different contacts are not necessarily correlated, the open-close transition may only take place if the fluctuations of the various contacts by chance cause the simultaneous switching of all relevant interactions. Hence we propose that a successful transition necessitates a concerted or cooperative action of all involved coordinates of the intramolecular contact network of T4L.

Single event analysis: essential coordinates. To facilitate a concise discussion of the structural changes associated with the open-closed transition, we need to further reduce the number of coordinates. In a first step, we disregard all contacts that do not change during the open-closed transition, but only fluctuate in either the contact-formed or contact-broken state. This leaves us with 18 hydrophobic contacts, 5 hydrogen bonds, and 4 salt bridges (Table S2). While hydrophobic contact changes typically involve complex rearrangements of various atoms that are difficult to grasp (see below), changes of hydrogen bonds and salt bridges can often be directly linked to overall structural changes and are therefore considered first.

Most prominently, we identify three salt bridges that change during the open-closed transition. That is, in the hinge region the salt bridge between Glu5 and Lys60 opens, while in the mouth region the salt bridges Asp20-Arg145 and Glu22-Arg137 close. The first salt bridge described by $d_{5,60}$ complements the previously defined locking coordinate $p \equiv d_{4,60}$, while the second salt bridge corresponds to the above discussed opening coordinate $x \equiv d_{20,145}$. Ultimately, the third salt bridge is associated with the contact distance $d_{22,137}$ describing the opening and closing of the tip of the mouth, see Fig. 2b. Apart from these fairly obvious interactions, we find one more salt bridge (between Arg8 and Glu64) and several hydrogen bonds that are located in the mouth region (described by $d_{21,105}$ and $d_{22,141}$) and in the hinge region ($d_{2,64}$, $d_{7,11}$, and $d_{7,12}$).

As a simple means to study the mutual relation of these coordinates, we considered their time evolution during individual open↔closed transitions of the trajectory (Fig. S2). Interestingly, we find that the salt bridge $d_{8,64}$ and the hydrogen bond $d_{2,64}$ compete with each other, facilitating a seesaw-like motion of helix 1 with respect to helix 3 (Fig. S3). Moreover, we note that hydrogen bond distances $d_{21,105}$ and $d_{22,141}$ show quite similar behavior as the main salt bridges and are therefore redundant. On the other hand, distance $d_{7,12}$ (or alternatively $d_{7,11}$) is found to exhibit diverse behavior, as it accounts for the (de)stabilization of helix 1 (see below). From the three side-chain dihedral angles of cluster 1 (χ_4, χ_{70}, and χ_{104}), χ_4 is most important, as it is found to directly account for the re-orientation of Phe4. While χ_{104} exhibits a similar behavior over time as χ_4, it shows clearly more fluctuations whose structural implications are resolved below. χ_{70} shows hardly additional information and is therefore neglected. Put together, this leaves us with coordinates $d_{5,60}$, $d_{4,60}$, χ_4 and $d_{7,11}$ reporting on the (un-)locking of Phe4 and the subsequent rearrangement of helix 1, as well as $d_{20,145}$ and $d_{22,137}$ describing the opening and closing of the mouth. Illustrating the mechanism of the transition, these coordinates can be

considered as essential internal coordinates.[30]

To illustrate to what extent these coordinates from the hinge to the mouth region indeed move simultaneously, we zoom into the time evolution of the system during a representative open→closed transition. Assigning the start of the transition to time $t = 0$, Fig. 3a reveals that distance $d_{5,60}$ changes first, before right afterwards coordinates $d_{4,60}$, $d_{22,137}$ and $d_{20,145}$ change virtually simultaneously. A few nanosecond later χ_4 responds, followed by $d_{7,12}$ about 10 ns later. Of particular interest is the simultaneous response of distances $d_{4,60}$ (hinge) and $d_{22,137}$ (mouth), which are far apart (about 2.5 nm) from each other (see Fig. 2b). This synchronism indeed seems to indicate a direct mechanical coupling of the two protein sections.

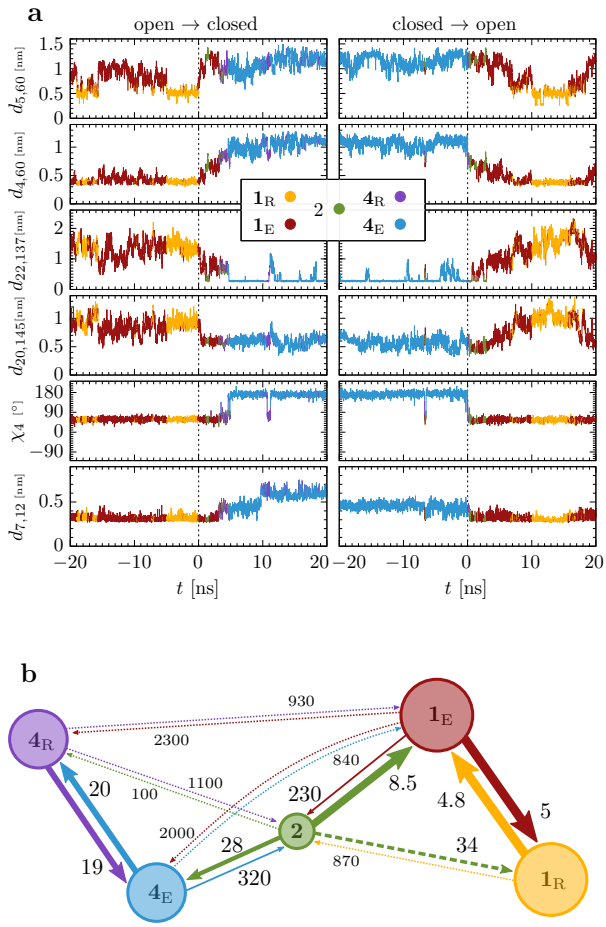

FIG. 3. (a) Time evolution of some key coordinates of T4L during a representative open→closed (left, occurring at 3.2 µs in Fig. 1) and closed→open (right, at 16.6 µs) transition, respectively. The color encodes the instantaneous conformational state of the system revealed by density-based clustering. (b) Scheme of the resulting Markov state model indicating the main transition times (in units of ns).

Structural characterization: transmission network. To explain the above findings, Fig. 4a depicts the overall conformational rearrangement of T4L during the open-closed transition. While the N- and C-terminal protein domains perform a mirrored rotational motion, helix 1 executes a piston-like move into the opposite direction. This structural coupling between N- and C-terminal domains as well as with helix 1 is facilitated by a network of persistent salt bridges indicated in Figs. 4b and S3. That is, helix 1 is connected to the C-terminal domain via several salt bridges between the two negatively charged residues Asp10 and Glu11 in helix 1 and the two positively charged residues Arg145 and Arg148 in helix 10 of the C-terminal domain. Along helix 3, we find a salt bridge between Asp72 and Arg76. Furthermore, helix 3 is connected to the N-terminal domain via Arg52-Glu62, and to the C-terminal domain via Arg80-Glu108. All these electrostatic connections serve as structural tethers that provide a balance between stability and flexibility, which we consider as a prerequisite for a bistable system. In Fig. 4b, panels (1) to (5) comprise the individual local sites mediating the conformational rearrangement from the protein open to the closed state. Moreover, the analysis suggests the intramolecular coordinates shown in Fig. 3a, which are indicative of the local structural changes. The structural evolution is illustrated in real time in the Supplementary Movie S1.

(1) To allow the exit of Phe4 from its hydrophobic pocket, first a salt bridge between Glu5 and Lys60 needs to open. This motion is naturally described by the contact distance $d_{5,60}$. The subsequent outward move of Phe4 can be monitored by the locking coordinate $d_{4,60}$.

(2) The outward motion of Phe4 is transferred via Phe67, Leu7 to Phe104 and its adjacent helix 5. That is, Phe4, Phe67, Leu7 and Phe104 form a hydrophobic core that causes a rolling motion of helix 5 over helix 1 similar to two cogwheels in contact. As helix 5 represents an integral part of the C-terminal domain, this motion of helix 5 rotates the C-terminal domain as shown in Fig. 4a. At the end of the rotation, Phe104 on the opposite end of helix 1 reacts via a change of its first side-chain dihedral angle χ_{104}.

(3) At the same time, the conformational change of the Phe67 side chain is transferred farther to the backbone of helix 3, which causes a straightening of the helix. This straightening acts on the N-terminal domain via the backbone and side chain packings, causing the forward movement of the N-terminal domain. In combination, the above described motions of the C- and the N-terminal domains lead to the closing of the mouth of T4L.

(4) The closed conformation is stabilized via a salt bridge at the tip of the mouth between Glu22 and Arg137, which can be followed via distance $d_{22,137}$, and an electrostatic interaction complex involving Glu11, Arg14, Asp20 and Arg145 that can be monitored by $d_{20,145}$ in the middle of the mouth. Both salt bridge centers impose structural near-order on

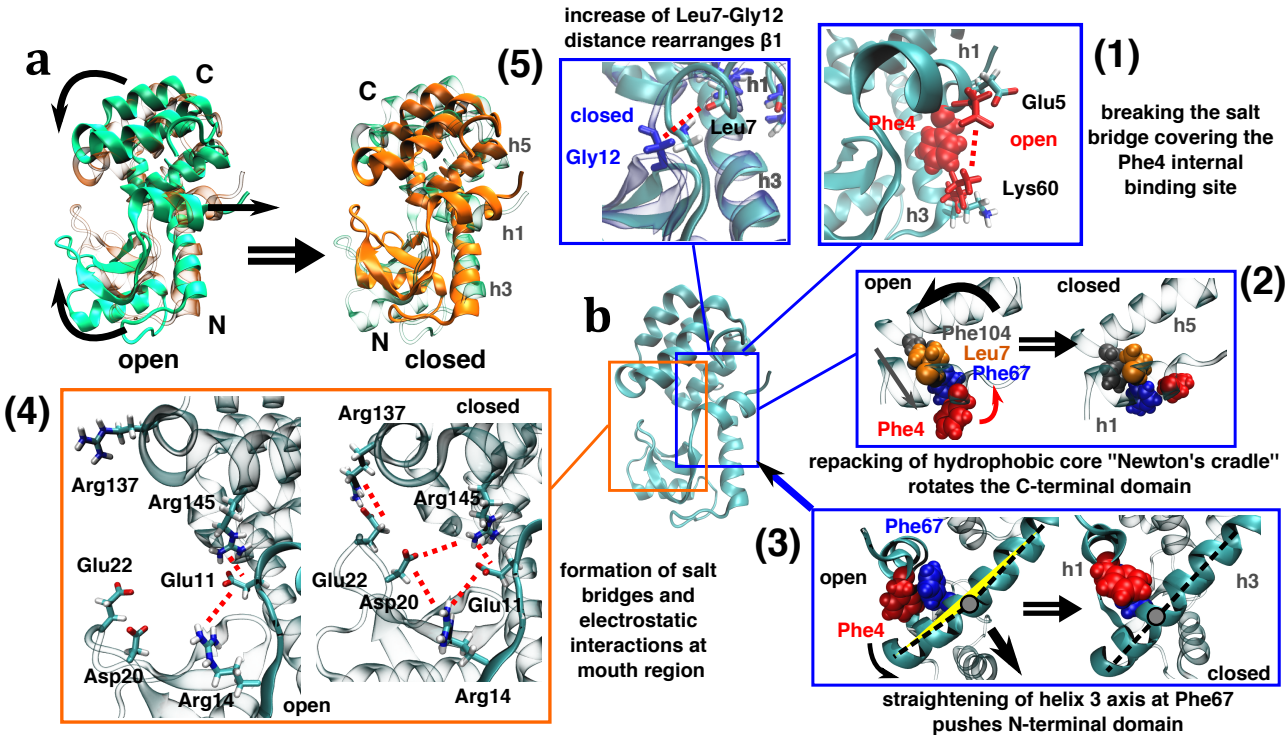

FIG. 4. Structural details of the open-closed mechanism in T4L, with black arrows indicating the direction of backbone motions. (a) illustrates the overall picture, (b) accounts for the individual steps of the conformational rearrangement, see the text.

the mouth region. Since these coordinates respond simultaneously with the locking coordinate $d_{4,60}$ (which indicates that the Phe4 side chain passes over the side chain of Phe67), the complex conformational rearrangements described in steps (2) to (4) occur in a concerted manner. Subsequently, the Phe4 side chain rotates into a new conformation and becomes completely exposed to the solvent. This can be monitored by the side-chain dihedral angle χ_4.

(5) Finally, a destabilization of the last winding of helix 1 leads to a rupture of a backbone hydrogen bond between Leu7 and Gly12, resulting in an increase in the distance $d_{7,12}$. As a consequence, the motion of Gly12 causes a change in the position of the β_1 strand such that Arg14 is rotated away from the electrostatic complex, thus weakening its coupling with Glu11 and Asp20.

To go back to the open state, the above steps simply occur in reverse order. Here the crucial issue is the locking of Phe4, i.e., its move from the solvent-exposed to a hydrophobically-buried state, which is indicated by coordinates $d_{7,12}$, χ_4 and p.

In summary, we have shown that the hydrophobic core (Phe4, Phe67, Leu7 and Phe104) mediating a cogwheel-like motion of helices 1 and 5 and the straightening of helix 3 constitutes a flexible transmission network, which represents the molecular machinery underlying the open-closed transition. While common motions of Phe4, Phe67 and Phe104 seemed plausible from the comparison of various wild-type and mutant crystal structures of T4L,[24,25] here we show that these residues together with Leu7 in fact perform a concerted motion. Since helices 3 and 5 are rigidly connected to the N- and C-terminal domains, respectively, the motion of the transmission network in the hinge region is directly and instantaneously coupled to the mouth region of T4L. This long-range or allosteric coupling is the origin of the concerted action of the 85 distances connecting the mouth and the hinge region of T4L (Fig. 2b).

State model description: heterogeneity of pathways. The finding of several steps in the cooperative mechanism raises the question to what extent these steps are associated with well-defined intermediate states of the process. To this end, we employed the six coordinates (x, p, $d_{5,60}$, $d_{22,137}$, χ_4, $d_{7,12}$) discussed in Fig. 3a and performed robust density-based clustering[36] in this space (see Material and methods for details). In line with the nomenclature of the previous four-state model (Fig. 1c), the resulting five conformational states are referred to as states $\mathbf{1}_\mathrm{R}$ (with a population of 33.5 %), $\mathbf{1}_\mathrm{E}$ (35.5 %), $\mathbf{2}$ (1.3 %), $\mathbf{4}_\mathrm{E}$ (15.1 %) and $\mathbf{4}_\mathrm{R}$ (14.5 %), see Fig. 3b. Compared to the previous model that relied on coordinates (x, p) only, notably the two main states $\mathbf{1}$ (open) and $\mathbf{4}$ (closed) each split up into two similarly populated substates, indexed by "R" (for relaxed) and "E" (for excited). In both cases, the sub-state E represents the "pole position" to start the open↔closed transition,

while the sub-state R accounts for the fully relaxed equilibrium state. Displaying the probability distributions of all six coordinates in each state, Fig. S4 reveals that the splitting of state **1** is caused by coordinate $d_{5,60}$, while the splitting of state **4** is caused by coordinate $d_{7,12}$. Although the intermediate state **2** is weakly populated and short-lived as in the previous model, it turns out to be mechanistically relevant (see below). In contrast, the weakly populated previous intermediate state **3** is now a part of state $\mathbf{4}_{\mathrm{E}}$. The spatial extension and the connectivity of the five states are well characterized by the two-dimensional free energy surfaces along the above defined six coordinates shown in Fig. S5.

To relate these conformational states to steps (1) - (5) in Fig. 4b, we consider the resulting state trajectory during the representative open→closed transition shown in Fig. 3a. As suggested above, we find that transition $\mathbf{1}_{\mathrm{R}} \rightarrow \mathbf{1}_{\mathrm{E}}$ accounts for the initial unlocking of Phe4 (Fig. 4b1). The subsequent main cooperative process (Figs. 4b2 - 4), on the other hand, is associated with two transitions, $\mathbf{1}_{\mathrm{E}} \rightarrow \mathbf{2}$ and $\mathbf{2} \rightarrow \mathbf{4}_{\mathrm{E}}$. State **2** represents the first stabilized closed structure, that is, it succeeded to stabilize one of the frequent attempts to close the mouth along opening coordinate x. Moreover, we find that helix 3 is already straightened in this state, reflecting step (3). Only by reaching state $\mathbf{4}_{\mathrm{E}}$, though, the cogwheel-like motion of the hydrophobic core and helices 3 and 5 (steps (2) and (3)) can be completed, with Phe4 moving into the solvent. Finally, transition $\mathbf{4}_{\mathrm{E}} \rightarrow \mathbf{4}_{\mathrm{R}}$ describes the stabilization of the closed state (step (5)), which typically takes longer than the nanosecond timescale shown in Fig. 3a.

To illustrate the heterogeneity of the open↔closed transitions and discuss the possible pathways of the reaction, Fig. S2 shows additional transition events apart from the one shown in Fig. 3a. Moreover, we calculated the resulting transition matrix $T(\tau_{\mathrm{lag}})$ containing the probabilities T_{ij} that the system jumps from state i to j within lag time τ_{lag}. To distinguish completed transitions from transient fluctuations, dynamical coring[37] was applied which requires that the trajectory spends a minimum time τ_{cor} in the new state for the transition to be counted (Fig. S6). Using $\tau_{\mathrm{cor}} = \tau_{\mathrm{lag}} = 0.7\,\mathrm{ns}$, the resulting transition times $t_{ij} = \tau_{\mathrm{lag}}/T_{ij}$ of the state model are shown in Fig. 3b. We note that the state doublets ($\mathbf{1}_{\mathrm{R}}$, $\mathbf{1}_{\mathrm{E}}$) and ($\mathbf{4}_{\mathrm{R}}$, $\mathbf{4}_{\mathrm{E}}$) interconvert frequently with 5 and 20 ns transition time, respectively. Apart from the main pathway $\mathbf{1}_{\mathrm{R}} \rightarrow \mathbf{1}_{\mathrm{E}} \rightarrow \mathbf{2} \rightarrow \mathbf{4}_{\mathrm{E}} \rightarrow \mathbf{4}_{\mathrm{R}}$ discussed above, we find various shortcuts of the reaction that, e.g., omit state $\mathbf{1}_{\mathrm{E}}$ (by going directly from $\mathbf{1}_{\mathrm{R}}$ to **2**) or omit state **2** (by going directly from $\mathbf{1}_{\mathrm{E}}$ to $\mathbf{4}_{\mathrm{E}}$), or omit state $\mathbf{4}_{\mathrm{E}}$ (by going directly from **2** to $\mathbf{4}_{\mathrm{R}}$).

When we attempt to construct a Markov state model from the transition matrix, we find that its predictions of the state-to-state waiting times are only in qualitative agreement with the corresponding MD results, i.e., they deviate from MD results by a factor of ∼2 (Fig. S6). This reflects the fact that during a successful cooperative open-close reaction the various state-to-state transitions occur almost simultaneously (Fig. 3a), i.e., they are not independent from each other.[38] This is opposed to the timescale separation assumed by a Markov state model, where the lag time (representing the randomization time of the fast intrastate dynamics) should be significantly shorter than the transition times of the model.

CONCLUSIONS

To investigate the microscopic mechanism underlying the open-closed transition of T4L, we have combined a correlation analysis of contact features (Fig. 2) with a study of their time evolution (Fig. 3) and a careful visualization of the conformational changes (Fig. 4 and Supplementary Movie S1). In particular, we have identified a set of essential internal coordinates (Fig. 3) that monitor various aspects of the transition. The allosteric mechanism in T4L is found to consist of three ingredients.

(i) To start the open-closed transition, a salt bridge between Glu5 and Lys60 needs to open, in order to allow the exit of Phe4 from its hydrophobic pocket (Fig. 4b1). The finalization of the overall transition requires a rearrangement of salt bridges in the mouth region (Fig. 4b4) and a hydrogen bond of helix 1 (Fig. 4b5), which both stabilize the closed state. In the case of the converse reaction, first the latter three interactions are reversed, while the final formation of the Glu5-Lys60 salt bridge stabilizes the closed state. That is, both open and closed states are stabilized by specific interactions, which initially need to be overcome to get the system in a pole position for the desired transition.

(ii) Following this initiation, the hydrophobic core (Phe4, Phe67, Leu7 and Phe104) of the hinge region mediates a cogwheel-like motion of helices 1 and 5 (Fig. 4b2) as well as the straightening of helix 3 (Fig. 4b3). In this way, the hydrophobic core constitute a flexible transmission network, which represents the molecular machine driving the open-closed transition.

(iii) Since helices 3 and 5 are tightly connected to the N- and C-terminal domain, respectively, the motion of the transmission network in the hinge region is directly and instantaneously coupled to the mouth region of T4L. Hence the small local changes of the hydrophobic core are passed on via rigid contacts and the protein backbone to distant sites.

We anticipate that these three components of the process are ubiquitous for allosteric transitions of bistable proteins. For example, a repacking of a hydrophobic core as well as switching of polar interaction partners was reported for the nitrogen regulatory protein C.[17] Moreover, the above described lever-arm effect is reminiscent of structural changes in G protein-coupled receptors,

where small changes in the ligand binding site are structurally amplified to large-scale protein surface changes at the G protein binding site.[39,40] A similar effect was also reported for the microbial rhodopsin bacteriorhodopsin, where the retinal cofactor pushing against Trp182 causes a large-scale outward motion of helix F,[41] and for the connection between ATP hydrolysis and protein conformational changes in heat shock protein 90.[20] The interplay between rigid secondary elements (such as α-helices and β-sheets) and flexible protein sections (such as loops or linkers) to mediate structural rearrangements was also discussed by Nussinov and Thirumalai.[9,42]

The simultaneous involvement of residues of the N- and C-terminal domains as well as of helix 1 demonstrates a high cooperativity of the open-closed transition. This finding is in line with recent unfolding experiments on T4L,[26] which revealed that helix 1 is responsible for the high degree of cooperativity within the protein. By coupling the two domains, helix 1 gives rise to an all-or-none, two-state unfolding behavior. It is interesting to note that the cooperativity of the (un)folding of T4L seems to transfer to the cooperativity of its functional dynamics. This insight could allow us to employ –at least to some extent– the well-established theories and concepts of protein folding to the less understood theoretical description of allosteric transitions. In this context we note that recent time-resolved experimental and computational studies of a photoswitchable PDZ2 domain indicated that allosteric communication shares some properties with multistate downhill folding.[8,19] Similar to the various degrees of cooperativity in protein folding,[43] allosteric transitions may as well show a variety of mechanisms.

Using the identified essential internal coordinates, density-based clustering revealed five conformational states (Fig. 3b), which showed a splitting of the two main open and closed protein states reflecting their stabilization described in (i). Apart from this and a weakly populated optional intermediate state, however, the resulting Markov state model provided only little insight in the cooperative open-close transition. This is because in a cooperative process the various motions occur almost simultaneously (Fig. 3a) and are therefore not independent from each other, which is opposed to the timescale separation assumed by a Markov model.

Cooperative processes such as two-state protein folding, on the other hand, have been successfully described by Ising-type models.[44,45] Here we consider an Ising model below the critical temperature, where local couplings between two-state entities lead to the emergence of well-defined global orientations of the system. Applied to the functional dynamics of a bistable protein, we associate the local two-state systems with the side-chain contacts, their next-neighbor couplings with interactions with residues in side-chain or backbone proximity, and the resulting global orientations with the open and closed protein conformations. In direct analogy to the system's stabilization in a single global orientation by an external field, in allostery a protein state can be stabilized via the (un)binding of a ligand.

The Ising model approach explains the cooperativity of the global open-closed transition via the interaction of local fluctuations.[46] While all local contacts fluctuate on their own, a successful conformational transition requires the simultaneous flipping of all contacts at the same time. In other words, the system needs to pass through a narrow region in phase space, which constitutes a rare event.[47] In particular, the need of a simultaneous change of numerous contacts explains the long waiting times despite small barriers along individual coordinates (Fig. 1). A similar combination of small barriers and transitions requiring maximal cooperativity between amino acids was found for the folding of small proteins.[48]

In conclusion, we have shown for a case study of T4L that the combination of extensive MD simulations, correlation-based dimensionality reduction, detailed inspection of the structural changes, and models from protein folding theory may provide a microscopic understanding of the allosteric mechanism in unprecedented detail. As in an Ising model, global conformational rearrangements and allosteric communication over long distances are found to emerge from local fluctuations and short-distance couplings.

MATERIAL AND METHODS

MD simulations This work is based on the 61 µs-long all-atom MD trajectory of Ernst et al.,[29] which used the GROMACS v2016 software package[49] and the Amber99*ILDN force field.[50–52]

Definition of coordinates The Leiden algorithm[35] is a community detection algorithm which identifies subparts of strongly interacting nodes in graphs. For the identification of collective motion in T4L we constructed a similarity-graph in which the nodes represent the coordinates and the edges indicate how correlated they are. For this graph, the Leiden algorithm maximizes an objective function, indicating to what extent the graph exhibits a clusters-like topology. Here we employed the constant Potts model as the objective function

$$\mathcal{H}_{\text{CPM}} = \sum_c \left[e_c - \gamma \binom{n_c}{2} \right], \tag{2}$$

where e_c denotes the sum of edge weights and n_c the number of nodes in cluster c. γ is a resolution parameter that governs the degree to which coordinates within a cluster must be correlated. Setting $\gamma = 0.5$, seven larger clusters were found, while the majority of the coordinates are mostly uncorrelated to the rest of the system. Cluster 1 corresponds to 85 coordinates, listed in Table S1, that correlate highly with the open-closed transition.

Characterization of metastable states Robust density-based clustering[36] identifies states based on a local free energy estimate which is calculated for each trajectory point by counting the number of neighbors found

inside a d-dimensional hypersphere of fixed radius R centered at the point. We applied this method to the coordinates $x = d_{20,145}$, $p = d_{4,60}$, $d_{5,60}$, $d_{22,137}$, χ_4 and $d_{7,12}$. To ensure that all six coordinates span similar value ranges the angle χ_4 was divided by 180. Using a clustering radius of $R = 0.034$ and a minimal population $P_{\min} = 100$ resulted in 24 state clusters. Of those 24 states, only the first 9 states have populations > 1 %. By inspecting the location of these states in the six-dimensional space, we were able to associate them with one of the 5 coarse-grained states $\mathbf{1}_{\mathrm{R}}$ (microstates 2 and 5), $\mathbf{1}_{\mathrm{E}}$ (microstate 1), $\mathbf{2}$ (microstate 9), $\mathbf{4}_{\mathrm{R}}$ (microstates 4 and 8) and $\mathbf{4}_{\mathrm{E}}$ (microstates 3, 6 and 7). The remaining 15 microstates were assigned to the closest coarse-grained state.

Markov state model We first applied dynamical coring to improve the state definition.[37] Here, short-living fluctuation are removed from the dynamics by requiring for valid transitions that the trajectory spends at least a minimum time τ_{cor} in the new state. To determine τ_{cor}, we considered the probability $P_{\mathrm{stay},i}(t)$ to stay in state i which should decay like a single exponential function. Short-lived fluctuation induce a fast initial decay, which can be removed by using $\tau_{\mathrm{cor}} = 700$ ps. To test the reliability of the model, we generated a Markov chain Monte Carlo trajectory and compared the average waiting times of MD and Markov model, i.e., the average times $t_{\mathrm{wait,i\to j}}$ needed to reach state j after entering state i (Fig. S6).

Acknowledgments

The authors thank Peter Hamm, Thorsten Hugel, Peter Bolhuis, Daniel Nagel, and Simon Stitzinger for fruitful discussions. This work has been supported by the Deutsche Forschungsgemeinschaft (DFG) via the Research Unit FOR 5099 "Reducing complexity of nonequilibrium" (project No. 431945604). The authors acknowledge support by the bwUniCluster computing initiative, the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC and the DFG through grant No. INST 37/935-1 FUGG.

* email: steffen.wolf@physik.uni-freiburg.de
† email: stock@physik.uni-freiburg.de

[1] K. Gunasekaran, B. Ma, and R. Nussinov, Is allostery an intrinsic property of all dynamic proteins?, Proteins **57**, 433 (2004).
[2] I. Bahar, C. Chennubhotla, and D. Tobi, Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation, Curr. Opin. Struct. Biol **17**, 633 (2007).
[3] Q. Cui and M. Karplus, Allostery and cooperativity revisited, Protein Science **17**, 1295 (2008).
[4] H. N. Motlagh, J. O. Wrabl, J. Li, and V. J. Hilser, The ensemble nature of allostery, Nature (London) **508**, 331 (2014).
[5] C. J. Tsai and R. Nussinov, A Unified View of "How Allostery Works", PLoS Comput. Biol. **10** (2014).
[6] E. Guarnera and I. N. Berezovsky, Allosteric sites: remote control in regulation of protein activity, Curr. Opin. Struct. Biol. **37**, 1 (2016).
[7] N. V. Dokholyan, Controlling allosteric networks in proteins, Chem. Rev. **116**, 6463 (2016).
[8] G. Stock and P. Hamm, A nonequilibrium approach to allosteric communication, Phil. Trans. B **373**, 20170187 (2018).
[9] D. Thirumalai, C. Hyeon, P. I. Zhuravlev, and G. H. Lorimer, Symmetry, rigidity, and allosteric signaling: From monomeric proteins to molecular machines, Chem. Rev. **119**, 6788 (2019).
[10] S. J. Wodak et al., Allostery in its many disguises: From theory to applications, Structure **27**, 566 (2019).
[11] M. Gruebele, Fast protein folding, in *Protein Folding, Misfolding and Aggregation*, edited by R. Schweitzer-Stenner, pages 106–138, The Royal Society of Chemistry, 2008.
[12] K. A. Dill, S. Banu Ozkan, M. Scott Shell, and T. R. Weikl, The protein folding problem, Annu. Rev. Biophys. **37**, 289 (2008).
[13] S. Brüschweiler, P. Schanda, K. Kloiber, B. Brutscher, G. Kontaxis, R. Konrat, and M. Tollinger, Direct observation of the dynamic process underlying allosteric signal transmission, J. Am. Chem. Soc. **131**, 3063 (2009).
[14] O. Bozovic, C. Zanobini, A. Gulzar, B. Jankovic, D. Buhrke, M. Post, S. Wolf, G. Stock, and P. Hamm, Real-time observation of ligand-induced allosteric transitions in a PDZ domain, Proc. Natl. Acad. Sci. USA **117**, 26031 (2020).
[15] O. Bozovic, J. Ruf, C. Zanobini, B. Jankovic, D. Buhrke, P. J. M. Johnson, and P. Hamm, The Speed of Allosteric Signaling Within a Single-Domain Protein., J. Phys. Chem. Lett. **12**, 4262 (2021).
[16] M. D. Vesper and B. L. de Groot, Collective dynamics underlying allosteric transitions in hemoglobin, PLoS Comp. Biol. **9**, e1003232 (2013).
[17] F. Pontiggia, D. Pachov, M. Clarkson, J. Villali, M. Hagan, V. Pande, and D. Kern, Free energy landscape of activation in a signalling protein at atomic resolution, Nat. Commun. **6**, 7284 (2015).
[18] C. A. Smith, D. Ban, S. Pratihar, K. Giller, M. Paulat, S. Becker, C. Griesinger, D. Lee, and B. L. de Groot, Allosteric switch regulates protein-protein binding through collective motion, Proc. Natl. Acad. Sci. USA **113**, 3269 (2016).
[19] S. Buchenberg, F. Sittel, and G. Stock, Time-resolved observation of protein allosteric communication, Proc. Natl. Acad. Sci. USA **114**, E6804 (2017).
[20] S. Wolf, B. Sohmen, B. Hellenkamp, J. Thurn, G. Stock, and T. Hugel, Hierarchical dynamics in allostery following ATP hydrolysis monitored by single molecule FRET measurements and MD simulations, Chem. Sci. **12**, 3350 (2021).

[21] M. Bhattacharyya, S. Ghosh, and S. Vishveshwara, Protein structure and function: Looking through the network of side-chain interactions, Curr. Prot. Pept. Sci. **17**, 4 (2016).

[22] G. R. Bowman, V. S. Pande, and F. Noé, *An Introduction to Markov State Models*, Springer, Heidelberg, 2013.

[23] U. Sengupta and B. Strodel, Markov models for the elucidation of allosteric regulation, Phil. Trans. R. Soc. B **373**, 20170178 (2018).

[24] M. Dixon, H. Nicholson, L. Shewchuk, W. Baase, and B. Matthews, Structure of a hinge-bending bacteriophage T4 lysozyme mutant, Ile3→Pro, J. Mol. Biol. **227**, 917 (1992).

[25] X. Zhang, J. A. Wozniak, and B. W. Matthews, Protein flexibility and adaptability seen in 25 crystal forms of T4 Lysozyme, J. Mol. Biol. **250**, 527 (1995).

[26] E. Shank, C. Cecconi, J. Dill, S. Marqusee, and C. Bustamante, The folding cooperativity of a protein is controlled by its chain topology, Nature (London) **465**, 637 (2010).

[27] H. Sanabria et al., Resolving dynamics and function of transient states in single enzyme molecules, Nat. Commun **11**, 1231 (2020).

[28] J. S. Hub and B. L. de Groot, Detection of functional modes in protein dynamics, PLoS Comput. Biol. **5**, e1000480 (2009).

[29] M. Ernst, S. Wolf, and G. Stock, Identification and validation of reaction coordinates describing protein functional motion: Hierarchical dynamics of T4 Lysozyme, J. Chem. Theory Comput. **13**, 5076 (2017).

[30] S. Brandt, F. Sittel, M. Ernst, and G. Stock, Machine learning of biomolecular reaction coordinates, J. Phys. Chem. Lett. **9**, 2144 (2018).

[31] F. Sittel and G. Stock, Perspective: Identification of collective coordinates and metastable states of protein dynamics, J. Chem. Phys. **149**, 150901 (2018).

[32] G. Diez, D. Nagel, and G. Stock, Correlation-based feature selection to identify functional dynamics in proteins, arXiv:2204.02770v2 (2022).

[33] R. B. Best and G. Hummer, Coordinate-dependent diffusion in protein folding, Proc. Natl. Acad. Sci. USA **107**, 1088 (2010).

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

[35] V. Traag, L. Waltman, and N. van Eck, From Louvain to Leiden: guaranteeing well-connected communities, Sci. Rep. **9**, 5233 (2019).

[36] F. Sittel and G. Stock, Robust density-based clustering to identify metastable conformational states of proteins, J. Chem. Theory Comput. **12**, 2426 (2016).

[37] D. Nagel, A. Weber, B. Lickert, and G. Stock, Dynamical coring of Markov state models, J. Chem. Phys. **150**, 094111 (2019).

[38] B. Lickert and G. Stock, Modeling non-Markovian data using Markov state and Langevin models, J. Chem. Phys. **153**, 244112 (2020).

[39] C.-M. Suomivuori et al., Molecular mechanism of biased signaling in a prototypical G protein–coupled receptor, Science **367**, 881 (2020).

[40] O. Fleetwood, J. Carlsson, and L. Delemotte, Identification of ligand-specific G protein-coupled receptor states and prediction of downstream efficacy via data-driven modeling, eLife **10**, e60715 (2021).

[41] T. Weinert et al., Proton uptake mechanism in bacteriorhodopsin captured by serial synchrotron crystallography, Science **365**, 61 (2019).

[42] E. Papaleo, G. Saladino, M. Lambrughi, K. Lindorff-Larsen, F. L. Gervasio, and R. Nussinov, The role of protein loops and linkers in conformational dynamics and allostery, Chem. Rev. **116**, 6391 (2016).

[43] P. Malhotra and J. B. Udgaonkar, How cooperative are protein folding and unfolding transitions?, Prot. Sci. **25**, 1924 (2016).

[44] E. R. Henry, R. B. Best, and W. A. Eaton, Comparing a simple theoretical model for protein folding with all-atom molecular dynamics simulations, Proc. Natl. Acad. Sci. USA **110**, 17880 (2013).

[45] S. Olsson and F. Noé, Dynamic graphical models of molecular kinetics, Proc. Natl. Acad. Sci. USA **116**, 15001 (2019).

[46] T. C. B. McLeish, T. L. Rodgers, and M. R. Wilson, Allostery without conformational change: modelling protein dynamics at multiple scales, Phys. Biol. **10**, 056004 (2013).

[47] C. Dellago and P. Bolhuis, Transition path sampling and other advanced simulation techniques for rare events, Adv. Polymer Sci. **221** (2016).

[48] S.-H. Chong and S. Ham, Time-dependent communication between multiple amino acids during protein folding, Chem. Sci. **12**, 5944 (2021).

[49] M. J. Abraham, T. Murtola, R. Schulz, S. Pall, J. C. Smith, B. Hess, and E. Lindahl, Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX **1**, 19 (2015).

[50] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, Proteins **65**, 712 (2006).

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

[52] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, Improved side-chain torsion potentials for the amber ff99sb protein force field., Proteins **78**, 1950 (2010).