Global motions exhibited by proteins in micro- to milliseconds simulations concur with anisotropic network model predictions

M. Gur,[a] E. Zomot,[b] and I. Bahar[c]
Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh,
3501 Fifth Ave, Suite 3064 BST3, Pittsburgh, Pennsylvania 15260, USA

The Anton supercomputing technology recently developed for efficient molecular dynamics simulations permits us to examine micro- to milli-second events at full atomic resolution for proteins in explicit water and lipid bilayer. It also permits us to investigate to what extent the collective motions predicted by network models (that have found broad use in molecular biophysics) agree with those exhibited by full-atomic long simulations. The present study focuses on Anton trajectories generated for two systems: the bovine pancreatic trypsin inhibitor, and an archaeal aspartate transporter, GltPh. The former, a thoroughly studied system, helps benchmark the method of comparative analysis, and the latter provides new insights into the mechanism of function of glutamate transporters. The principal modes of motion derived from both simulations closely overlap with those predicted for each system by the anisotropic network model (ANM). Notably, the ANM modes define the collective mechanisms, or the pathways on conformational energy landscape, that underlie the passage between the crystal structure and substates visited in simulations. In particular, the lowest frequency ANM modes facilitate the conversion between the most probable substates, lending support to the view that easy access to functional substates is a robust determinant of evolutionarily selected native contact topology. © *2013 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported License.* [http://dx.doi.org/10.1063/1.4816375]

I. INTRODUCTION

Proteins undergo continual conformational changes under physiological conditions while maintaining their overall fold. These changes vary from local fluctuations to large scale domain/subunit movements. Local changes, typically over timescales of up to tens of nanoseconds, can be efficiently modeled by all-atom simulations; whereas global transitions, which usually occur over microseconds or slower, are largely beyond the capacity of full-atomic simulations. The latter class of motions is usually of interest due to its direct relevance to biological function. In view of these issues, analytical approaches such as normal mode analysis (NMA) in conjunction with coarse-grained models, such as the Anisotropic Network Model (ANM)[1] (shortly termed ANM analysis), have found broad applicability for exploring, albeit at low resolution, the global changes accessible to biomolecular systems.

ANM is a simple physics-based model of beads and springs, which exclusively depends on inter-residue contact topology. Despite its simplicity, several studies have shown that (i) the global motions (lowest frequency modes) predicted by the ANM are practically indistinguishable from those obtained by NMA with full-fledged atomic force fields, and (ii) these modes often relate to functional changes in structure, such as the fluctuations between the unbound (open) and bound (closed) conformers of a given enzyme, or the passage between the different substates of allosteric proteins.[2–5] However, except for the work of Grossfield and co-workers on the dynamics of G-protein coupled receptors,[6,7] a direct comparison of ANM-predicted collective motions with that observed in all-atom micro-to-milliseconds molecular dynamics (MD) simulations has been elusive due to the prohibitively expensive task of generating such long MD trajectories. The recently developed supercomputing machine Anton[8,9] now permits us to generate MD trajectories that may serve as a test bed for examining, or establishing, the utility of ANM predictions. Here, we analyze two Anton-generated trajectories, one longer than a millisecond on the equilibrium dynamics of bovine pancreatic trypsin inhibitor (BPTI),[10] and the other, of several microseconds, on the gating mechanism of a transporter, the archaeal aspartate transporter GltPh[11] that serves as a model for human excitatory amino acid transporters.[12–15] Our analysis clearly shows that the ANM successfully captures the global dynamics of both systems obtained by these extensive MD simulations, while these simulations themselves may have their own limitations.

BPTI is a widely studied small protein, with several high-resolution crystal structures resolved since 1975.[16] It served as the subject macromolecule of the first reported MD simulation,[17] and since then its dynamics have been extensively studied by MD,[18–23] NMA,[24–28] and experiments.[29–33] We will use its recent Anton trajectory of 1.013 ms, which proved to reproduce numerous experimental findings,[10] as a benchmark for establishing the method for analyzing the

---
[a]E-mail: meg78@pitt.edu
[b]E-mail: enz1@pitt.edu
[c]Author to whom correspondence should be addressed. Electronic mail: bahar@pitt.edu. Telephone: (412) 648-3333. Fax: (412) 648-3163.

output from simulations and comparing the extracted information with ANM predictions. A close similarity between the first three principal components (PCs), or essential motions, obtained by principal component analysis (PCA) of Anton trajectory and the three lowest frequency ANM modes will be shown. Notably, the free energy surface constructed for BPTI will demonstrate that the passages between the most probable substates are essentially driven by the structure-encoded softest ANM modes.

GltPh is the only structurally characterized member of the biologically important family of excitatory amino acid transporters. Even though various conformations of GltPh have been resolved to date,[15,34–36] and several simulations have been performed using these structural data,[11,37–43] the mechanism of function of aspartate transporters (and their mammalian glutamate orthologs) is not yet well-understood. GltPh is a homotrimer. Each of its monomers are composed of two domains: a "scaffold" (transmembrane (TM) helices 1, 2, 4, and 5) that provides support for the transport core and forms the interface between the monomers, and a transport core (TM3, 6, 7, 8; and two helical hairpins, HP1 and HP2) that binds and translocates the substrate and co-transported sodium ions. Substrate transport is enabled by alternating access to the extracellular (EC) and intracellular (IC) environments, which is accomplished by the global transitions of GltPh between its outward-facing (OF) and inward-facing (IF) states, during substrate uptake and release, respectively. Experimental and computational studies have pointed to the role of two hairpins, HP1 and HP2, in gating the substrate-binding site.[11,15,35,36,39,42,44] The conformations of HP1 and HP2 will be shown below to be important determinants of the functional substates sampled during two Anton trajectories of 6 μs each, generated for the IF state of GltPh. Strikingly, in line with our findings on BPTI, the essential motions extracted from the Anton simulations for GltPh in the IF state show close overlap with low frequency ANM modes. Moreover, these ANM modes appear to naturally facilitate the transitions between the substates sampled in the simulations.

II. MATERIALS AND METHODS

A. MD simulations

The BPTI run was performed in explicit water under NVT conditions (T = 300 K) as described in detail in Shaw *et al.*[10] The protein and ions were modeled with the AMBER ff99SB force field,[45] and water molecules were modeled with the 4-particle TIP4P-Ew force field.[46] GltPh simulations were carried out under NPT conditions, at 310 K and 1 bar (for details, see Zomot and Bahar[11]). CHARMM27 force field parameters with CMAP terms were used for the protein and water molecules, and CHARMM36 parameters, for lipid molecules.[47,48] The system was composed of 121 000 atoms solvated in 150 mM NaCl, including ~30 000 explicit water molecules (TIP3P) and 365 1-palmitoyl-2-oleoyl-phosphatidyl ethanolamine (POPE) lipid molecules with explicit head groups and united-atoms for the tail groups. The two Anton runs performed for IF GltPh will be referred to as

MD1 and MD2. Each comprises two parts, designated as holo (before aspartate release) and apo (after aspartate release).

B. Principal component analysis

The covariance matrix **C** is constructed using[49]

$$\mathbf{C} = \langle (\mathbf{R} - \langle \mathbf{R} \rangle)(\mathbf{R} - \langle \mathbf{R} \rangle)^{\mathrm{T}} \rangle, \qquad (1)$$

where **R** is the 3N-dimensional configuration vector composed of the instantaneous C^α-atom coordinates of the N residues of the protein, and $\langle \mathbf{R} \rangle$ is its average over the trajectory. PCs are obtained by eigenvalue decomposition of **C** as

$$\mathbf{C} = \sum_{i=1}^{3N} \sigma_i \mathbf{p}_i \mathbf{p}_i^T. \qquad (2)$$

Here $\mathbf{p}_i$ is the ith PC (ith eigenvector) and σ_i is the corresponding variance (eigenvalue), ordered in descending order with respect to σ_i. σ_i scales with the magnitude of the motion along $\mathbf{p}_i$. The first PC, $\mathbf{p}_1$, and its variance, σ_1, define the shape and size of the lowest frequency motion of the protein.

C. Anisotropic network model

In the ANM,[1] a potential of the form,

$$V_{\mathrm{ANM}} = -\frac{\gamma}{2} \left[\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(R_{ij} - R_{ij}^0 \right)^2 \Gamma_{ij} \right], \qquad (3)$$

is used to construct the Hessian matrix, **H**,[50] where R_{ij}^0 and R_{ij} are the original (native state) and instantaneous distances between the C^α-atoms, i and j, and Γ_{ij} is the ijth element of the Kirchhoff matrix **Γ**, equal to -1 if $R_{ij}^0 < r_c$, and zero otherwise. Residue pairs with C^α–C^α distance smaller than cut-off distance r_c are assumed to interact ($r_c = 12$ and 15 Å for GltPh and BPTI, respectively). Eigenvalue decomposition of **H** yields $3N - 6$ nonzero eigenvalues λ_k ordered as $\lambda_1 \le \lambda_2 \le \cdots < \lambda_{3N-6}$, and the corresponding eigenvectors $\mathbf{u}_k^T = \left[u_k^{X1}, u_k^{Y1}, u_k^{Z1}, \ldots, u_k^{ZN} \right]$. The kth eigenvector (counterpart of $\mathbf{p}_k$) describes the normalized displacements of the N residues along the kth mode, and $1/\lambda_k$ (counterpart of PCA σ_k) scales the size of these displacements (although the order of the modes may change between PCA and ANM analysis).

D. Overlap between ANM modes and PCs

Overlap between ANM modes and PCs is evaluated using the correlation cosine between the eigenvectors,[51,52]

$$\mathrm{I}_{i,j} = |\mathbf{u}_i \cdot \mathbf{p}_j|, \qquad (4)$$

and the cumulative overlap between the first k ANM modes and jth PC is defined as[50]

$$CO_{k,j} = \left(\sum_{i=1}^{k} I_{i,j}^2 \right)^{\frac{1}{2}}. \qquad (5)$$

E. Deformation vector

The unit deformation vector between substates l and m is defined as

$$\mathbf{d}_{l,m} = \frac{\mathbf{R}^{(m)} - \mathbf{R}^{(l)}}{\|\mathbf{R}^{(m)} - \mathbf{R}^{(l)}\|}, \qquad (6)$$

where $\mathbf{R}^{(m)}$ is the configuration vector of substate m. The ability of mode k to assist in the passage from substate l to substate m is measured by the correlation cosine, $O = \mathbf{u}_k \cdot \mathbf{d}_{l,m}$.

F. Free energy calculations

The entropy of a given substate is evaluated based on the number of associated microstates Ω, as $S = k \ln \Omega$. To this aim, the conformational space sampled by MD is divided into grids, which are suitably grouped into subspaces. If $N(\Phi_s)$ designates the number of conformers in each of these subspaces, s, the number of microstates can be directly related to the number of conformers in each subspace, $\Omega \approx N(\Phi_s)$. Assuming that the energy of each substate is approximately the same, the free energy surface is evaluated using the probability distribution function $f(\mathbf{R}) = N(\Phi_s)/\sum_s N(\Phi_s)$ in $A(\mathbf{R})$ $= -kT \ln \{f(\mathbf{R})\} + \mathrm{ct}$.

III. RESULTS AND DISCUSSION

A. Bovine pancreatic trypsin inhibitor dynamics

1. BPTI conformations sampled in Anton simulations map onto eight substates

Figure 1 provides an overview of the distribution of the substates deduced from the PCA of the 1.013 ms long Anton trajectory generated[10] for native-state BPTI. Panels (a) and (b) display the projections of conformers onto the subspace spanned by three PCs. The PCs are selected therein so as to allow for distinctive visualization of the conformational subspaces occupied by the eight substates. Figure S1 in the supplementary material[53] provides a detailed description of the time evolution of the instantaneous C^α-atoms-based root-mean-square deviation (RMSD) from the BPTI crystal structure (Protein Data Bank (PDB) code 5PTI[54]) (panel (a)), and the evolutions along the top-ranking five principal components (panels (b)–(f)) during this run. The histograms along the right ordinate in each case display the distribution of accessible regions, which is multimodal along particular principal axes, overall leading to the eight substates S1–S8 displayed in Figure 1. For example, the fluctuations projected onto PC1 (Figure S1(b) of the supplementary material[53]) help discern S1, S2, and S5 from the others; S1, S2, and S5, themselves, can be differentiated along PC2 (Figure S1(c) of the supplementary material[53]); and the projections onto PC4 and PC5 separate S6 and S7 from others (Figures S1(e)–S1(f) of the supplementary material[53]). Substate S2 is the closest (RMSD of ~1 Å) to the crystallographic structure (Figure 1(c)) and is the second most populated substate, after S1, in accord with the kinetic clustering by Shaw *et al.*[10] Table SI of the supplementary material lists the RMSD values between all pairs.[53] Figure S2 of the supplementary material[53]
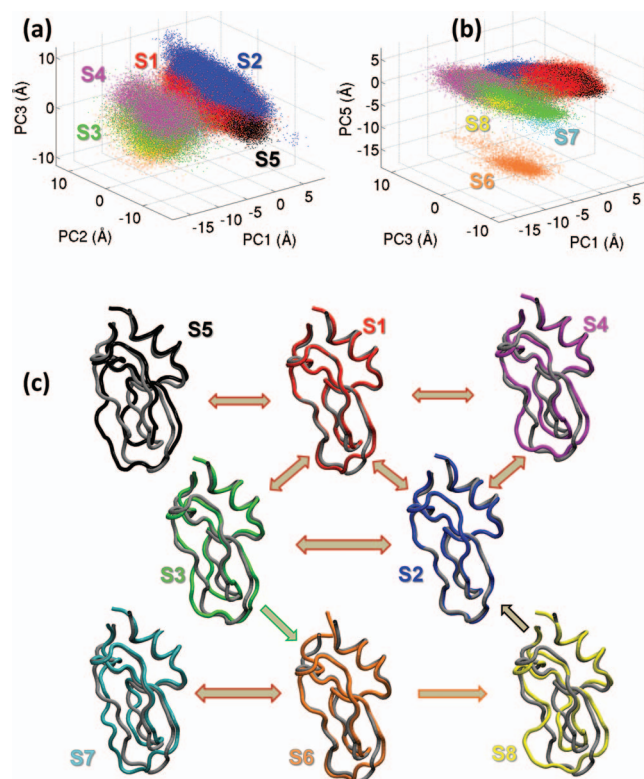

FIG. 1. Distribution of BPTI substates sampled during MD simulations. (a) Distribution of conformers projected onto the subspace spanned by the three principal components PC1–PC3 obtained from PCA of the 1.013 ms trajectory. (b) Same as in (a) projected onto the subspace spanned by PC1, PC3, and PC5. Occupancy percentage of states and coloring are as follows: Substate 1 (S1; 55.3%, *red*), substate 2 (S2; 26.6%, *blue*), substate 3 (S3; 6.6%, green), substate 4 (S4; 5.5%, *magenta*), substate 5 (S5; 2.4%, *black*), substate 6 (S6; 1.2%, *orange*), substate 7 (S7; 0.3%, *cyan*), and substate 8 (S8; 0.24%, *yellow*). (c) Structures representative of substates S1–S8 (*colored*) superimposed onto the crystal structure (*grey*). Transitions observed during the course of simulations are indicated by arrows.

displays their inter-residue contact maps, with the contacts formed/broken with respect to substate S2 highlighted in *green/red*.

The populations (or fractional occurrences) of these substates are given in Figure 1 caption, and the accessible transitions and observed first passage times/rates are presented in Table SII of the supplementary material.[53] Note that out of the $N_s(N_s - 1) = 56$ types of transitions possible for a system of $N_s = 8$ substates, we observe 17 only, and many of them are visited only once. Therefore, the tabulated first passage times (varying between 4.0 and 80 μs) are approximate values presented here to provide an estimate of the time scales of observed events, and their inverses do not represent the transition rates or probabilities between these pairs. In principle, under equilibrium conditions, the inter-conversions between substates a and b should obey detailed balance, $p^o(a)\, p(b|a) = p^o(b)p(a|b)$, where $p^o(a)$ is the equilibrium probability of a, and $p(b|a)$ is the conditional/transition probability from b to a. Among the observed transitions, S1 ↔ S2 conforms (approximately) to the detailed balance principle (Table SII of the supplementary material[53]). This transition involves the most probable two substates (which account for 81.9% of the conformational space, with $p^o(\text{S1}) = 0.553$ and $p^o(\text{S2}) = 0.266$;

see Figure 1 caption). The departure of other passages from detailed balance indicates sampling inaccuracies that are incurred due to low probabilities of the associated conformers, or rare events of inter-conversions, even in these milliseconds simulations of equilibrium dynamics for a small protein.

2. Global motions extracted from MD trajectories are similar to those predicted by the ANM

Eigenvalue decomposition of the $3N \times 3N$ covariance matrix **C**, derived from the BPTI Anton trajectory, yields $3N - 6 = 168$ non-zero PCs ($N = 58$ residues). The heat map in Figure 2(a) shows the overlap between the first five PCs (PC1–PC5) and the five lowest frequency modes predicted for the crystal structure by the ANM (ANM1–ANM5). The top-ranking three modes show good agreement between the two sets. In particular, PC1 yields a correlation of 0.70 with ANM2. Figure 2(b) compares the square displacement profiles of amino acids along PC1 and along ANM2, suggesting that the first principal mode extracted from this expensive simulation could be readily predicted (to a good approximation) by ANM analysis of the starting structure.

In order to understand how well the slowest ANM modes capture the principal motions, we evaluated the corresponding cumulative overlaps (Eq. (5)), shown in the *right* column in Figure 2(a). The resulting high overlaps with PC1, PC2, and PC3 further corroborate the consistency between the principal directions of motion deduced from the MD trajectory and the top-ranking ANM modes based exclusively on the inter-residue contact topology (of the initial structure).

Finally, the covariance matrices constructed using the three lowest-frequency ANM modes (Figure 2(c)) and the top-ranking three PCs (Figure 2(d)) exhibit similar patterns. The most distinctive cross-correlations are those between the segment Pro13-Ala16 (covariance indices 37–48) and other residues, including the N-terminus, Cys38-Arg39 and Gly57-Ala58. Two cysteines in these regions, Cys14 and 38, form one of the three disulfide bonds that stabilize the tertiary structure of BPTI (Figure S3 of the supplementary material[53]).

3. ANM modes favor transitions between most probable substates of BPTI

Next, we investigated the ability of the ANM modes to account for the conformational transitions between the crystal structure (represented by S2) and other substates. For this purpose, $3N$-dimensional deformation directional vectors $\mathbf{d}_{l,m}$ (Eq. (6)) were constructed for the starting substate S2 ($l = 2$) and various ending substates ($1 \leq m \leq 8$), and the correlation cosines between $\mathbf{d}_{l,m}$ and the most probable ANM modes $\mathbf{u}_k$ ($1 \leq k \leq 5$) were examined (see Sec. II).

As illustrated in Figures 3(b) and 3(d), ANM1 (or $\mathbf{u}_1$) exhibits a strong correlation with the deformation vector $\mathbf{d}_{2,1}$ between the two most probable substates, S1 and S2: this single mode (among the set of $3N - 6$ ANM modes) accounts alone for 57% of the transition. Figure 3(a) shows how the direction of the collective motion driven by ANM1 (shown by *purple arrows*) computed for S2 conforms to the reconfiguration of S2 (*blue*) into S1 (*red*).
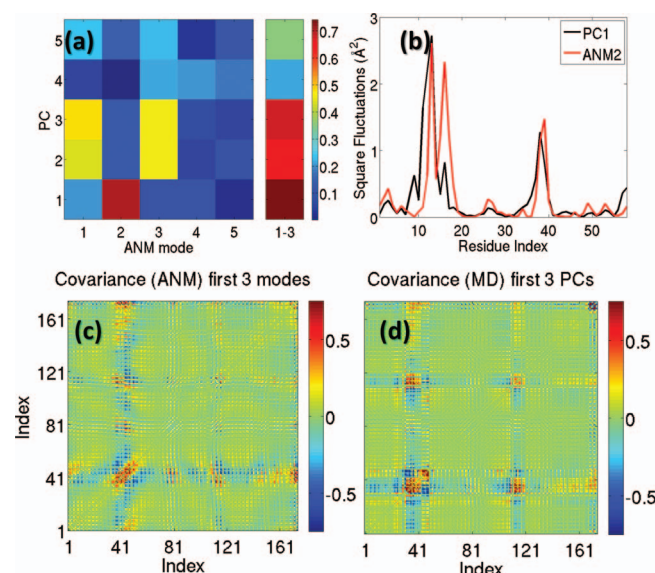

FIG. 2. Correlations between principal modes (PCs) observed in MD and low frequency modes predicted by the ANM. (a) The map displays the overlap between PC1–PC5 and ANM1–ANM5. The column on the *right* shows the cumulative overlap between ANM1–ANM3 and each PC. (b) Comparison of square displacement profiles of C$^\alpha$-atoms along the PC1 (*black*) and ANM2 (*red*). ANM2 profile is rescaled so that it shares the same norm as that of PC1. (c) Covariance matrix predicted ANM1-ANM3. The indices along the two axes correspond to the x-, y-, and z-components of the N residues. (d) Covariance matrix based on PC1-PC3. Diagonal elements are set to zero to enhance the contrast of off-diagonal elements.
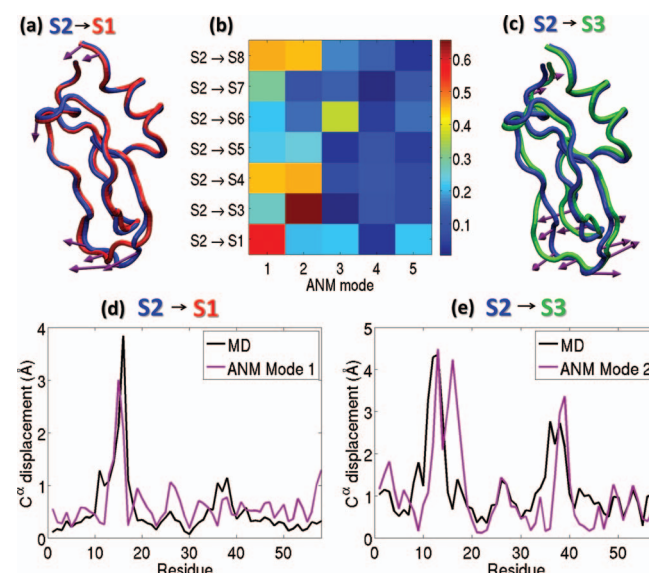
FIG. 3. Contribution of ANM modes from the crystal structure (represented by S2) to all other substates. (a) The structural change from S2 (*blue*) to S1 (*red*) concurs with the direction of motion defined by ANM1 (*purple arrows*). (b) Correlation cosines between the five lowest frequency ANM modes (abscissa) and the deformation vector corresponding to the transitions listed along the ordinate. (c) Reconfiguration of S2 (*blue*) into S3 (*green*) along ANM2 (*purple*). (d) Comparison of change in C$^\alpha$-coordinates between S2 and S1 (*black*), and the ANM1 predicted for S2 (*purple*). ANM1 profile is uniformly rescaled to match the norm of the displacement vector. (e) Same as in (d), for the passage between S2 and S3 in comparison to ANM2.
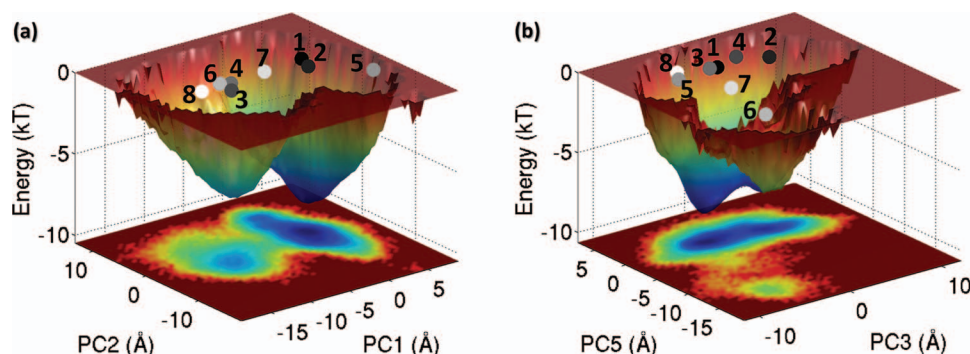
FIG. 4. Free energy surface based on principal modes of motion. Free energy surface projected onto (a) PC1 and PC2, and (b) PC3 and PC5. The surfaces are expressed relative to their lowest energy in units of kT. Substates are shown as beads and colored on a white-grey-black scale based on their index, S1 being black and S8 white.

Systematic analysis of the transitions from the crystal structure (represented by S2) to all observed substates showed that ANM2 essentially drives the transition from S2 to the third most populated substate S3 (Figures 3(c) and 3(e)) with a correlation cosine of 0.66 (Figure 3(b)). We also note that ANM1 and ANM2 are jointly involved in the transitions to S4 and S8 (cumulative overlaps of ∼0.64 in each case). Transition of S2 into S5–S7 yielded relatively small correlation cosines with ANM1–ANM5, consistent with the observation that in the complete MD trajectory no such transitions were observed (Figure S1 and Table SII of the supplementary material[53]).

The lowest frequency modes take place, by definition, along the directions of smallest ascent away from the original energy minimum. The energy along the slowest (and all modes) has a quadratic shape, with curvature (or stiffness) scaling with the eigenvalue of the mode. The curvature is smallest in the case of lowest energy modes, hence their description as "soft" modes. In the present case, the passage between the substates S1 and S2 is enabled by the softest ANM mode, ANM1, and the overall movement between the two endpoints is less than 1 Å RMSD. This small movement along this soft mode would therefore involve a small energy increase, or a low energy barrier, as is observed in the free energy surface displayed in Figure 3. The fact that the two most populated substates, S1 and S2, interconvert with the help of ANM1 is noteworthy, attesting to the significance of the slowest ANM modes in facilitating, if not enabling, the most relevant transitions. Furthermore, the third most populated substate, S3, is accessed via ANM2, further supporting the significance of the lowest frequency modes predicted by the ANM as mechanisms facilitating the excursion on the conformational energy landscape.

4. The two most populated BPTI substates, S1 and S2, are separated by a low energy barrier

Figure 4(a) illustrates the free energy landscape for BPTI constructed based on the 1.013 ms Anton trajectory in the subspace spanned by PC1 and PC2 (see Sec. II). Two broad minima are observed, one comprising the substates S1, S2, and S5, and the other comprising S3, S4, S6, and S8. These two minima are separated by a large energy barrier with S7 located near the saddle point. Figure 4(b) displays the free energy surface in the subspace spanned by PC3 and PC5. Substates S1–S5 and S8 merge into a broad minimum in this case, separated from S6, with S7 again appearing near the saddle point. We note that S1, S2, and S5 co-localize in the same minimum in both subspaces.

Further investigation of the energy landscapes in other subspaces (Figure S4 of the supplementary material[53]) showed that some substates are separated by low energy barriers, such as S1 and S2, whereas others do not readily allow for interconversions due to high energy barriers, such as S1 and S6. S1 and S2 are separated mostly along PC3 and there is no energy barrier between them in the subspace spanned by PC1 and PC2. PC3 shows its highest correlation with ANM1 (Figure 2(a)), and ANM1 is the most readily accessible mode, which explains its emergence as the path of passage between S1 and S2 in Figure 3. It is interesting to note that the major structural difference between S1 and S2 takes place at the trypsin binding site near K15 and A16 (at the lower portion of the ribbon diagram in Figure 3(a)). Examination of available crystal structures for the bound and unbound forms of BPTI showed changes in side chain orientation at that particular site and in the position of the C-terminal tail, in accord with ANM1 predictions. The comparison of available crystal structures can be seen in Figure S5 of the supplementary material.[53]

B. Dynamics of aspartate transporter GltPh in the inward-facing state

1. Five distinctive substates are sampled during Anton simulations

We collected an ensemble of 12 000 equally spaced snapshots from Anton simulations (two runs of 6 μs each) performed[11] for GltPh homotrimer in the IF state, starting from the holo (with bound aspartate and two sodium ions, termed Na1 and Na2) crystal structure (PDB ID: 3KBC[36]). PCA of the trajectories based on TM helical C^α-coordinates
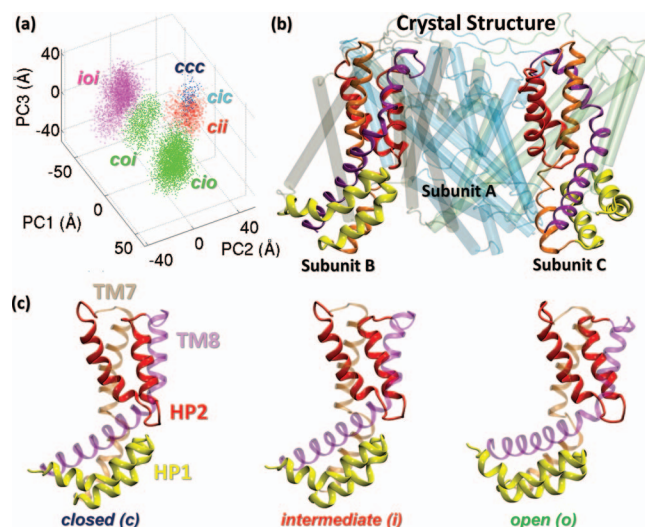
FIG. 5. Substates sampled in microseconds Anton trajectories generated for the IF state of GltPh. Panel (a) displays the distribution of the trimeric substates in the subspace spanned by PC1-PC3. Panel (b) shows the original crystal structure in the IF state. Transport core elements are color-coded (HP2, *red*; HP1, *yellow*; TM7, *orange*; TM8, *violet*) in two of the subunits. Panel (c) illustrates three conformers, open (*o*), intermediate (*i*), and closed (*c*), observed for individual subunits. Their selected combinations define the trimeric substates shown in panel (a). Conformers are distinguished by the relative positions of HP1 and HP2 tips (Figure S6 of the supplementary material[53]) and their RMSD from the crystal structure (Figure 6).

yielded the clusters shown in Figure 5(a), upon projecting the conformers onto the subspace spanned by PC1–PC3.

In order to characterize the conformations in each cluster, we used three metrics: the distance between the tips of HP1 and HP2 (using C^α-atoms of Ser277 and Val355, respectively; Figure S6 of the supplementary material[53]) in each subunit, the RMSDs of HP1 (and HP2) C^α-atoms from their crystal coordinates (Figure 6(a)), and the RMSDs in the inter-residue distances between HP1 (and HP2) and their neighbors within 12 Å (Figure 6(b)). These variables helped us unambiguously identify three conformers, open (*o*), intermediate (*i*), and closed (*c*) for each subunit, mainly distinguished by the conformations of transport core elements HP1, HP2, TM7, and TM8 (see Figure 5(c)).

Because GltPh is a cylindrically symmetric homotrimer, the identity of the subunit (A, B, or C) that is in a given conformer (*c*, *i*, or *o*) is immaterial. Therefore, the number of distinctive substates potentially accessible to the trimer would theoretically be ten: *ccc, cci, cco, cii, cio, coo, iii, iio, ioo, ooo*, if each subunit conformer was independent. Notably, five are observed out of these ten, in the simulations (Figure 6 and Figure S6 of the supplementary material[53]): (1) all three subunits in the closed conformer (*ccc, blue,* percent occurrence: 1.2%), (2) two subunits closed, and one intermediate (*cci; cyan,* 0.8%), (3) one closed, and two intermediate (*cii; red,* 6.7%), (4) each subunit in a different conformer (*cio* or *coi; green,* 53.5%), and finally (5) two intermediate and one open subunit (*ioi; magenta,* 33.2%). The RMSD values between these substates are listed in Table SIII of the supplementary material,[53] showing structural changes up to 3.4 Å. The most commonly observed substate (*cio, green*) contains two clusters, where the *o* and *i* conformers are

swapped between subunits B and C (using subunit designations in the PDB). Substates *ccc, cic,* and *cii* are co-localized in Figure 5(a), but they separate into distinct regions when projected onto higher PCs. Substate *ccc* is the closest to the crystallographic structure (RMSD of ~1.3 Å).

2. No simultaneous opening of two or more subunits is observed within microseconds

Examination of the accessible substates showed that no full opening of more than one subunit was observed within the time frame of simulations (i.e., substates *ooi, ooc,* or *ooo* were not sampled). Furthermore, we have not observed a simultaneous (partial) opening of the three subunits (i.e., substate *iii*), nor substate *cco*. The most "open" substate was that composed of one open and two intermediate subunits (*ioi* in MD2; Figures 5 and 6, and Figure S6 of the supplementary material[53]) accessed after the release of aspartate and sodium ions. These observations suggest an inter-subunit coupling that restricts simultaneous opening of subunits. This coupling is reminiscent of the observations made for GltPh simulations in the OF state,[42] which were attributed to intersubunit steric effects restraining the simultaneous opening of all three HP2 gates. The ability of the transport cores of the individual subunits to come into close proximity was confirmed experimentally for the OF state of GltPh.[41] The present results suggest that such intersubunit couplings apply to the IF state as well.

Our Anton simulations showed that a displacement of at least 4 Å in HP2 loop was a prerequisite for initiating aspartate release.[11] This displacement was invariably preceded by the release of Na2, and accompanied after a short time lag by HP1 opening.[11] HP1 movements would become even larger than those of HP2 *following* substrate release as can be seen in Figure 6 (the light/dark colors along the upper bars indicate the holo/apo portions of the two 6 μs runs). The high mobility of HP1 is presumably due to the disruption of the tight network of interactions at the substrate binding pocket succeeding substrate release. It is conceivable that these HP1 movements, which permit for an increased exposure of the binding site to the IC region, play a role in facilitating the release of the more deeply buried sodium ion Na1, an event that was not observed (apart from a single incidence) within the time scale of the simulations.

3. The closed trimer (ccc) tends to rapidly transition into the asymmetric intermediate coi/cio

GltPh simulations exhibit the progress of conformational events, from the opening of the EC gate to the release of sodium ion (Na2) and then aspartate, succeeded by the fluctuations in the neighborhood of the apo IF state; these runs do not reflect the equilibrium dynamics near a well-defined state, but a time evolution.[11] In line with the nature of the observed events, not all substates communicate with each other. Table SIV of the supplementary material[53] lists all observed transitions and associated first passage times, which vary by 2–3 orders of magnitude. The observed transitions exhibit the kinetic scheme *ccc* → *cic* → *cii* → *coi* → *ioi*, although the
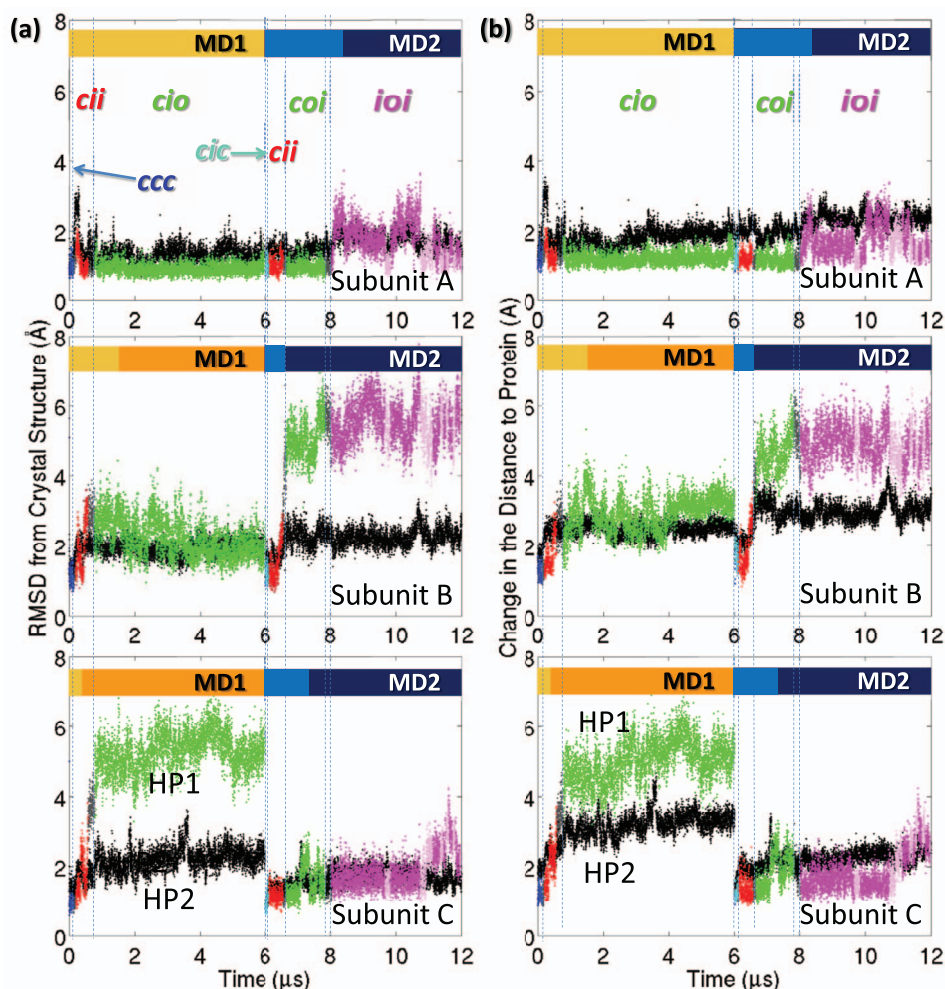
FIG. 6. Time evolution of HP1 and HP2 motions in the IF state of GltPh. Results for HP1 (colored by trimeric substates; see Figure S6 of the supplementary material[53]) and HP2 (*black*) are presented for each subunit over the entire course of runs MD1 (0–6 μs) and MD2 (6—12 μs). (a) RMSDs (Å) in HP1 and HP2 C^α-atoms with respect to their crystal structure positions. (b) RMSDs in inter-residue distances between HP1 and all other residues within 12 Å cutoff distance (*colored*) and between HP2 and all others (*black*). Distances are based on C^α-coordinates. The upper bars delimit the holo (*lighter color*) and apo (*darker color, after aspartate release*) portions of the two MD simulations. Aspartate was released in five out of six observations (two runs with three subunits each), except for subunit A in MD1.

intermediates *ccc*, *cic*, and *cii* are short-lived, and the dominant transition is *ccc* → *cio* (in MD1) which further progresses as *ccc* → *coi* → *ioi* (in MD2). The system therefore evolves from the closed form, to an asymmetric intermediate where on subunit's transport core fully opens (*o*), the 2nd partially opens (*i*), and the third remains closed (*c*), while on a global scale, the structure remains in the vicinity of IF state.

4. GltPh essential dynamics and passages between populated substates correlate with ANM modes

As shown in our earlier study, ANM1 and ANM2 are doubly degenerate modes that induce asymmetric openings of the three subunits (yielding a mixture of *c*, *o*, and *i* conformers for the overall subunits); whereas ANM3, the first non-degenerate mode, induces a concerted symmetric opening/closing of the three subunits.[38,41] The time scale of present simulations is too short for observing such concerted movements. However, it is interesting to note that the

PCA of the Anton trajectory indicates principal movements that exhibit some similarities with respect to ANM global modes.

The heat map in Figure 7(a) displays the overlap between the softest ANM modes evaluated for the initial IF trimer and the top-ranking PCs derived from simulations (note that ANM global modes exhibit little dependence on the selected substate; Figure S7 of the supplementary material[53]). PC1 yields an overlap of 0.71 with ANM1 and ANM5 (combined) (Figure 7(a)). The mobility profiles predicted by PC1 and ANM1 exhibit similar patterns (Figure 7(b)). The variance σ_1 of PC1 is 7.6 and 8.2 times larger than the respective variances of PC2 and PC3, making PC1 the most prominent dynamical feature in the trajectory. PC3 and ANM2 exhibit an overlap of 0.89 (Figure 7); and PC4 matches ANM3 mode (overlap of 0.73) (Figure S8 of the supplementary material[53] and Figure 7(a)). In addition to these pairwise correlations, we note the large cumulative overlaps between each of the top-ranking PCs and ANM1–ANM10 (*right* column in Figure 7(a)), which further substantiate the correlation
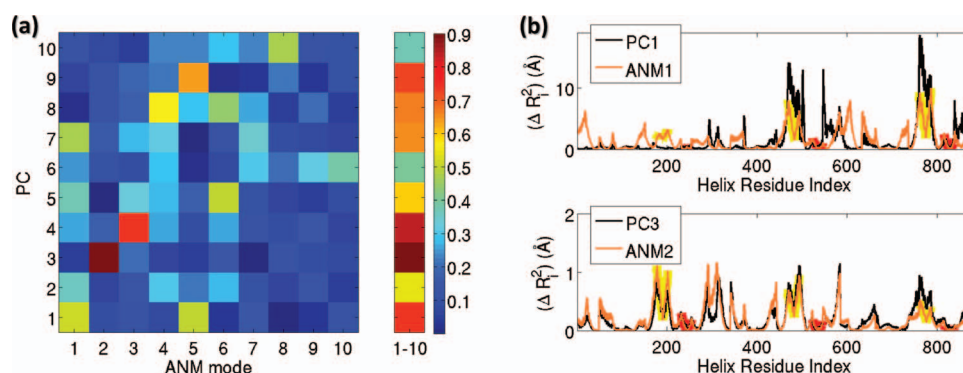
FIG. 7. Comparison of the top PCA modes sampled in the IF GltPh simulations and the lowest frequency ANM modes predicted for the trimer. (a) Overlap between PC1-PC10 from simulations and ANM1-ANM10 based on the initial coordinates of helical residues. The last column displays the cumulative overlap between each PC and ANM1-ANM10. (b) Distribution of square displacements of helical residues (C^{α}-atoms; 876 residues, 292/subunit) driven by the selected PC (*black*) and ANM (*orange*) modes (labeled) (see also Figure S7 of the supplementary material[53]). The HP1 regions are highlighted in *yellow*, and HP2 in *red*, in each subunit, both panels. ANM profiles are uniformly rescaled with respect to those of PC. Abscissa numbering starts with Val12 (subunit A) and ends with Thr415 (subunit C).

between ANM predictions and essential motions of GltPh observed in Anton simulations.

During the $ccc \rightarrow cii \rightarrow cio$ transition of the trimer, the HP1-HP2 separation increases from 4.9 Å to ~15 Å (Figure S6 of the supplementary material[53] and Figure 6), fully exposing the binding site (of subunit C in MD1, subunit B in MD2) to the IC space. Likewise, the $ccc \rightarrow cci \rightarrow cii \rightarrow coi \rightarrow ioi$ transitions (MD2 in Figure S6 of the supplementary material[53] and Figure 6) lead to an RMSD of about 15.0 Å in the HP1 domain of subunit B. Comparison of the deformation vectors associated with these structural changes with PC and ANM modes (Figure S9 of the supplementary material[53]) shows that the modes ANM1/PC1 (identified to be the counterpart of each other) facilitate the transition to cio; modes ANM2/PC3 drive the transition to coi and then ioi. Finally, if one focuses on the cooperative passage from ccc to cio (or coi), we can see in panel (c) of Figure S9 of the supplementary material[53] that this asymmetric change in structure correlates with the slowest doubly degenerate ANM modes (ANM1 and ANM2).

IV. CONCLUSION

In this study, we presented detailed analyses of the dynamics and conformational transitions of BPTI and GltPh in comparison to the collective motions predicted by the ANM. ANM is a simple structure and physics-based model, and requires only inter-residue contact information. Two major drawbacks could have limited the level of agreement between the two sets: the sampling inadequacy in MD, and coarse-graining of structure and energetics in the ANM. Despite these two potential sources of discrepancy, the global modes of motions predicted by the ANM exhibited reasonable correlations with the essential modes deduced by PCA of MD trajectories. The MD simulations indicated several structurally distinct substates and conformational changes associated with transitions between these substates, in accord with the reconfigurations that would be predicted away from the starting structure using the ANM analysis exclusively. These results support the use of the ANM as a useful tool for a first ex-

ploration of accessible conformational space as well as the most likely directions of reconfiguration in the neighborhood of a given state. Our study demonstrates that ANM lowest frequency modes essentially assist in the passage between the most probable substates, and suggests that proteins (or native inter-residue contact topologies) have evolved to favor functional changes in their structure.

Clearly, the MD results for BPTI are more exhaustive, while those for GltPh serve as investigational data. In this respect, it was interesting to observe that the free energy landscape of BPTI showed that conformational transitions along low frequency ANM modes incurred small energy barriers, in accord with the ANM theory. However, we also noted that despite the durations of the order of microseconds (GltPh) to milliseconds (BPTI) being attained by Anton simulations, the trajectories were short of a complete, statistically rigorous mapping of the conformational space (even under native state conditions). Notably, many substates were not adequately sampled even in BPTI simulations, and the transitions leading to, or away from, them would not obey detailed balance principle, and would not lend themselves to quantitative evaluation of transition rates. These limitations of MD simulations, despite the capability of Anton technology to access processes 1–2 orders of magnitude longer than those investigated by conventional software, further highlight the significance of adopting low resolution models and methods for guiding or complementing full atomic simulations, using for example the ANM,[55–58] in order to sample functional events and/or extract robust features relevant to biological function.

The events observed in GltPh simulations were confined to even shorter timescales, due to the shorter duration of simulations. Essentially, we observed transitions between open, intermediate, and closed states of the hairpins that control IC gating, which are of the order of tenths of microseconds or faster. Nevertheless, the Anton simulations were able to highlight long-range inter-subunit couplings that restricted these *local* motions, thus drawing attention to the coupling between local and global changes in structure. It was remarkable to see that careful examination of these apparently local (IC gate opening) events to extract the underlying essential

motions gave us hints about collective motions of the overall trimer. These observations support the view that MD simulations usually hide information on potential reconfigurations that would take place at longer-time or larger-length scales.[56] Multiple runs and rigorous analyses by PCA-based tools appear to be essential for extracting cooperative/global movements robustly favored by the native contact topology.

ACKNOWLEDGMENTS

[1] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, Biophys. J. **80**, 505 (2001).
[2] I. Bahar, T. R. Lezon, L. W. Yang, and E. Eyal, Annu. Rev. Biophys. **39**, 23 (2010).
[3] A. Bakan and I. Bahar, Proc. Natl. Acad. Sci. U.S.A. **106**, 14349 (2009).
[4] D. Tobi and I. Bahar, Proc. Natl. Acad. Sci. U.S.A. **102**, 18908 (2005).
[5] O. Keskin, BMC Struct. Biol. **7**, 31 (2007).
[6] A. Grossfield, S. E. Feller, and M. C. Pitman, Proteins **67**, 31 (2007).
[7] T. D. Romo and A. Grossfield, Proteins **79**, 23 (2011).
[8] R. O. Dror, M. O. Jensen, and D. E. Shaw, Conf. Proc. IEEE Eng. Med. Biol. Soc. **2009**, 2340 (2009).
[9] J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw, Curr. Opin. Struct. Biol. **19**, 120 (2009).
[10] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. B. Shan, and W. Wriggers, Science **330**, 341 (2010).
[11] E. Zomot and I. Bahar, J. Biol. Chem. **288**, 8231 (2013).
[12] G. E. Torres and S. G. Amara, Curr. Opin. Neurobiol. **17**, 304 (2007).
[13] S. G. Amara, Nature (London) **360**, 420 (1992).
[14] E. Zomot, A. Bakan, I. H. Shrivastava, J. DeChancie, T. R. Lezon, and I. Bahar, *Sodium-coupled Secondary Transporters, Molecular Machines* (World Scientific Publishing Co. Pte. Ltd., 2011), p. 199.
[15] D. Yernool, O. Boudker, Y. Jin, and E. Gouaux, Nature (London) **431**, 811 (2004).
[16] M. Karplus and J. A. McCammon, Nat. Struct. Biol. **9**, 646 (2002).
[17] J. A. McCammon, B. R. Gelin, and M. Karplus, Nature (London) **267**, 585 (1977).
[18] D. van der Spoel, A. R. van Buuren, D. P. Tieleman, and H. J. Berendsen, J. Biomol. NMR **8**, 229 (1996).
[19] S. Swaminathan, T. Ichiye, W. Van Gunsteren, and M. Karplus, Biochemistry **21**, 5230 (1982).
[20] M. Levitt, Nature (London) **294**, 379 (1981).
[21] M. Levitt and R. Sharon, Proc. Natl. Acad. Sci. U.S.A. **85**, 7557 (1988).
[22] R. M. Brunne and W. F. van Gunsteren, FEBS Lett. **323**, 215 (1993).
[23] C. L. Brooks III, A. Brunger, and M. Karplus, Biopolymers **24**, 843 (1985).
[24] B. Brooks and M. Karplus, Proc. Natl. Acad. Sci. U.S.A **80**, 6571 (1983).
[25] N. Go, T. Noguti, and T. Nishikawa, Proc. Natl. Acad. Sci. U.S.A. **80**, 3696 (1983).
[26] S. Hayward, A. Kitao, and N. Go, Protein Sci. **3**, 936 (1994).
[27] M. Levitt, C. Sander, and P. S. Stern, J. Mol. Biol. **181**, 423 (1985).
[28] T. Nishikawa and N. Go, Proteins **2**, 308 (1987).
[29] M. S. Appavou, G. Gibrat, and M. C. Bellissent-Funel, Biochim. Biophys. Acta **1764**, 414 (2006).
[30] S. Cusack, J. Smith, J. Finney, B. Tidor, and M. Karplus, J. Mol. Biol. **202**, 903 (1988).
[31] K. Wuthrich and G. Wagner, FEBS Lett. **50**, 265 (1975).
[32] G. Wagner, D. Bruhwiler, and K. Wuthrich, J. Mol. Biol. **196**, 227 (1987).
[33] G. Otting, E. Liepinsh, and K. Wuthrich, Biochemistry **32**, 3571 (1993).
[34] G. Verdon and O. Boudker, Nat. Struct. Mol. Biol. **19**, 355 (2012).
[35] O. Boudker, R. M. Ryan, D. Yernool, K. Shimamoto, and E. Gouaux, Nature (London) **445**, 387 (2007).
[36] N. Reyes, C. Ginter, and O. Boudker, Nature (London) **462**, 880 (2009).
[37] J. DeChancie, I. H. Shrivastava, and I. Bahar, Mol. Biosyst. **7**, 832 (2011).
[38] T. R. Lezon and I. Bahar, Biophys. J. **102**, 1331 (2012).
[39] Z. Huang and E. Tajkhorshid, Biophys. J. **95**, 2292 (2008).
[40] S. Stolzenberg, G. Khelashvili, and H. Weinstein, J. Phys. Chem. B **116**, 5372 (2012).
[41] J. Jiang, I. H. Shrivastava, S. D. Watts, I. Bahar, and S. G. Amara, Proc. Natl. Acad. Sci. U.S.A. **108**, 15141 (2011).
[42] I. H. Shrivastava, J. Jiang, S. G. Amara, and I. Bahar, J. Biol. Chem. **283**, 28680 (2008).
[43] Y. Gu, I. H. Shrivastava, S. G. Amara, and I. Bahar, Proc. Natl. Acad. Sci. U.S.A. **106**, 2589 (2009).
[44] B. H. Leighton, R. P. Seal, S. D. Watts, M. O. Skyba, and S. G. Amara, J. Biol. Chem. **281**, 29788 (2006).
[45] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Proteins **65**, 712 (2006).
[46] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon, J. Chem. Phys. **120**, 9665 (2004).
[47] R. W. Pastor and A. D. MacKerell, Jr., J. Phys. Chem. Lett. **2**, 1526 (2011).
[48] O. Guvench and A. D. MacKerell, Jr., Methods Mol. Biol. **443**, 63 (2008).
[49] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen, Proteins **17**, 412 (1993).
[50] Q. Cui and I. Bahar, *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems* (Chapman & Hall, 2005).
[51] F. Tama and Y. H. Sanejouand, Protein Eng. **14**, 1 (2001).
[52] L. Yang, G. Song, A. Carriquiry, and R. L. Jernigan, Structure (London) **16**, 321 (2008).
[53] See supplementary material at http://dx.doi.org/10.1063/1.4816375 for additional figures and tables.
[54] A. Wlodawer, J. Walter, R. Huber, and L. Sjolin, J. Mol. Biol. **180**, 301 (1984).
[55] B. Isin, K. Schulten, E. Tajkhorshid, and I. Bahar, Biophys. J. **95**, 789 (2008).
[56] L. Liu, A. M. Gronenborn, and I. Bahar, Proteins **80**, 616 (2011).
[57] Z. Yang, P. Majek, and I. Bahar, PLoS Comput. Biol. **5**, e1000360 (2009).
[58] Y. Arkun and M. Gur, PLoS ONE **7**, e29628 (2012).