JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ARTICLE

pubs.acs.org/JACS

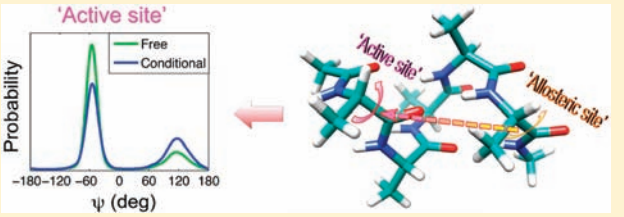
Atomistic Kinetic Model for Population Shift and Allostery in Biomolecules

Dong Long and Rafael Brüschweiler*

Department of Chemistry and Biochemistry and National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32306, United States

Supporting Information

ABSTRACT: Allosteric signaling in biomolecules is a key mechanism for a myriad of cellular processes. We present a general yet compact model for protein allostery at atomic detail to quantitatively explain and predict structural-dynamics properties of allosteric signal propagation. The master equation-based approach for allostery by population shift (MAPS) is introduced that derives the time scales, amplitudes, and pathways of signal transmission in peptides and proteins from dihedral angle dynamics observed in extended molecular dynamics simulations. The MAPS approach is first applied to the alanine-pentapeptide, and the results are tested against an explicit simulation in the presence of local conformational constraints, confirming the validity and accuracy of the model. We then apply the approach to a larger Markovian system based on a millisecond all-atom protein molecular dynamics trajectory of BPTI (Shaw et al. *Science* **2010**, *330*, 341−346). We use MAPS to illustrate in silico the propagation of a local perturbation over medium- to long-range distances across a disulfide bridge linking loops L1 and L2, which constitute the binding interface of BPTI.



INTRODUCTION

Allosteric regulation plays a key role in a myriad of biomacromolecular processes in living systems from gene expression to metabolism.[1,2] Specifically, the process of allostery refers to the transmission of a local perturbation, such as the binding of a ligand or a covalent modification,[3] on a protein to a distant site that is beyond the direct influence of the original perturbation,[4] and it has been argued that such an allosteric property is intrinsic to most proteins.[5] While the importance of allostery is widely recognized,[6−9] its physical mechanism is not well understood. The traditional interpretation of the allosteric mechanism has been based on either the concerted model[10] or the sequential model.[11] More recently, allostery[2,5,12−14] has been viewed in terms of protein conformational ensembles, which reflect the thermal fluctuations of protein structures, where signal transmission is explained by population shifts of conformational substates[5,12,15−17] induced by perturbations. While the ensemble view is sufficiently general to accommodate a wide variety of allosteric phenomena, the complexity associated with the adequate representation and analysis of a potentially large number of conformational substates has limited its predictive power. The development of a compact model for understanding allostery at atomic detail, which is physically and mathematically rigorous, is therefore an important objective.

The capability of transmitting a local structural-dynamic perturbation to a distant site relies on the intrinsic properties of dynamical coupling within a biomolecular system. In theory, a long equilibrium trajectory of an unperturbed system, for example, a protein in its apo state, statistically sampling all relevant conformations involved in an allosteric transition is sufficient for the prediction of the amplitudes, pathways, and time scales of signal transmission upon a given perturbation, provided that it follows a population shift mechanism. Recently, analysis of intrinsic motional coupling in dihedral angle space in the model protein ubiquitin revealed the dominance of short-range correlations with remarkably few medium- or long-range correlations.[18−20] Ubiquitin is a notably stable protein, which efficiently samples its functional substates separated by low energy barriers within its native energy basin on the submicrosecond time scale.[21] In the present study, we investigate the allosteric propensities of biomolecular systems that exhibit conformational transitions between multiple distinct states separated by higher energy barriers of several k_BT.

In an allosteric system, the initial perturbation alters the distribution of the degrees of freedom (DOF, e.g., dihedral angles) describing a local allosteric region (also known as the effector region), which is in direct contact with the ligand. In the presence of dynamical coupling, the altered distribution of the DOF of the allosteric region will affect the distributions of the DOF of the rest of the protein, which amounts to signal transmission. The resulting conditional ensemble, representing the probability distribution of the DOF of the rest region in the presence of the constraint probability distribution of the DOF of the allosteric region, contains all essential information about the structural-dynamic coupling between the different protein parts. The representation of protein molecules as Markovian systems[22−26] was demonstrated recently to be a suitable approach for capturing their kinetic

Received: September 19, 2011
Published: October 24, 2011

and thermodynamic properties, suggesting a means to quantitatively describe protein allostery. By extracting from an MD trajectory the kinetics of the transition network involving the various protein substates, it is possible to construct a Markov model capable of describing the populations of individual substates and their time evolution starting from an arbitrary initial population distribution. Importantly, the Markov model, represented by a master equation, can then be subjected to constraints on populations and conformational transitions, which after numerical integration of the modified master equation permits the systematic investigation of perturbations and their propagation within the biomolecule. This allows direct analysis of intramolecular communication among multiple distant sites.

In this work, we implement this concept using a master equation approach[24] based on molecular kinetics in dihedral angle space, capable of describing both the amplitude and the time evolution of signal transmission within a biomolecule in a quantitative and predictive manner. This approach is illustrated for a pentapeptide permitting a thorough in silico analysis of signal transmission via the population shift mechanism. In a second study, the model is further explored by application to the significantly larger protein system Bovine Pancreatic Trypsin Inhibitor (BPTI) based on a recent atomistic millisecond time scale molecular dynamics (MD) trajectory.[27]

METHODS

Molecular Dynamics Simulation. A 2 μs simulation of the capped alanine pentapeptide, (CH_3CO-Ala_5-$NHCH_3$; denoted hereafter as Ala_5) was performed with the AMBER-GSS force field,[28] ported to GROMACS,[29] and the explicit TIP3P solvent model[30] using the GROMACS software package (version 4.0.7).[31] This force field was shown to provide good agreement between simulated and experimental helix−coil equilibria.[32] Cut-off distances for the van der Waals interactions and short-range electrostatic interactions were set to 10 and 8 Å, respectively, and the long-range electrostatic interactions were treated with the particle-mesh Ewald summation method.[33] All bonds involving hydrogen atoms were constrained using the LINCS algorithm,[34] and the integration time step was set to 2 fs. A fully helical conformation of the peptide was used as starting conformation. Standard minimization and equilibration procedures described previously[21] were applied prior to the final production run at 300 K. Conformations were stored every 1 ps, yielding an ensemble with 2×10^6 conformations.

For BPTI, an all-atom 1.03 ms trajectory[27] by D. E. Shaw Research was used in the present study. The trajectory was run at 300 K under the NVT condition using a variant of the Amber ff99SB force field[35] incorporating modified isoleucine parameters[36] and the TIP4P-Ew water model.[37] A total of 41 250 conformations were used for subsequent analysis sampled every 25 ns.

Master Equation. For simplicity, the master equation approach is explained in the following for the pentapeptide, but its generalization to other systems, including BPTI, is straightforward. To construct a discrete master equation, the conformation of each individual residue of the polyalanine peptide is assigned to either a helical ("α") or nonhelical (coil) ("β") state based on its backbone φ,ψ dihedral angles using the transition-path-based assignment.[24] Two circular regions (with radius of 15°) centered around $(\varphi,\psi) = (-60°,-50°)$ and $(-60°,120°)$ were defined as helical and coil, respectively. Only the transitions from one region to another were recognized as change of states to suppress fast non-Markovian dynamics (such as fast librational motions). Similar to previous work,[24] the pentapeptide Ala_5 is described by a total of $2^5 = 32$ states, ranging from $\beta\beta\beta\beta\beta$ (all coil) to $\alpha\alpha\alpha\alpha\alpha$ (all helix); the corresponding decimal numbering of the states goes from 1 (all coil) to 32 (all helix). The transition rate constant (k_{ji}) from state i to state j is calculated as $k_{ji} = (N_{ij} + N_{ji})/(2T_i)$ and $k_{ii} = -\sum_{j\neq i} k_{ji}$, where N_{ji} is the number of transitions from i to j and T_i is the total residence time of state i.

The master equation can be expressed by a set of first-order linear differential equations:

$$\mathrm{d}\mathbf{P}/\mathrm{d}t = \mathbf{K}\cdot\mathbf{P}(t) \tag{1}$$

where $\mathbf{K}$ is the 32×32 transition rate constant matrix with elements k_{ji}, and $\mathbf{P}(t) = (p_1, p_2, ..., p_{32})^T$ is a column vector that contains the 32 populations of the individual substates fulfilling $\sum_{k=1}^{32} p_k = 1$, $p_k \geq 0$. Although the rate constant matrix $\mathbf{K}$ is derived from an equilibrium MD simulation, the master equation is capable of capturing the nonequilibrium dynamics (relaxation) of a pre-steady state, for example induced by a ligand-binding event, toward the new equilibrium. Importantly, the equilibrium properties of $\mathbf{K}$ can be biased to study the propagation of a perturbation through the protein system as described in the following section.

Master Equation-Based Approach for Allostery by Population Shift (MAPS). The effect of an external perturbation on a local residue can be described as a bias in the kinetic rates. Without loss of generality, we investigate the situation that the C-terminal residue Ala5 is restrained to the coil state (β) after perturbation. Such effects can be modeled by preferentially biasing the XXXXβ manifold over the XXXXα manifold of conformational substates, where X represents an arbitrary conformation β or α, through selective rescaling of the rate constants:

$$k_{ji}^{(\mathrm{new})} = \varepsilon\cdot k_{ji}^{(\mathrm{old})} \text{ and } k_{ij}^{(\mathrm{new})} = \frac{1}{\varepsilon}\cdot k_{ij}^{(\mathrm{old})} \tag{2}$$

where i and j represent any state from the XXXXβ and XXXXα manifolds, respectively, and ε is a prefactor ($0 < \varepsilon < 1$), which is set to 10^{-3} in this study, representing a strong local perturbation of the system. Under this constraint, the new equilibrium populations are directly given by

$$\mathbf{P}(t \rightarrow \infty) = |0\rangle \tag{3}$$

where $|0\rangle$ is the eigenvector with zero eigenvalue of the modified $\mathbf{K}$ matrix satisfying $\sum_{j=1}^{32} |0\rangle_j = 1$. It is noteworthy that the new equilibrium is solely determined by the modified kinetic matrix, independent of the initial condition $\mathbf{P}(0)$. The change of the coil population of residue S of Ala_5 is then given by the sum over all elements j of the vector $\Delta\mathbf{P} = \mathbf{P}(t\rightarrow\infty) - \mathbf{P}(0)$ in which residue S is in the coil state. For example, for $S = 2$, $\Delta P_{S=2} = \sum_{j\in \mathrm{X}\beta\mathrm{XXX}} \Delta P_j(t \rightarrow \infty)$ (where the sum goes over the 2^4 conformations with X (=α or β) of residues 1, 3, 4, and 5).

To investigate the role of individual residues after initial perturbation, the conformational transition ($\alpha \leftrightarrow \beta$) of a residue S can be selectively switched off by setting the rate constants $k_{nm} = k_{nm} = 0$, where m and n represent all substates where residue S is not in the same conformation. Only those transitions are retained in $\mathbf{K}$ that leave the conformation of residue S unchanged, and, hence, residue S is prevented from participating in any (allosteric) transmission pathway. Such constraints can be simultaneously applied to multiple residues to suppress their effect on signal transmission. In this manner, a subset of states can be selectively "decoupled" from the rest, which results in a rate constant matrix with more than one zero eigenvalue. To account for this situation, eq 3 needs to be modified as follows:

$$\mathbf{P}(t \rightarrow \infty) = \sum_{j}^{\lambda_j = 0} |v\rangle_j \langle w|_j \mathbf{P}(0) \tag{4}$$

where λ_j are the eigenvalues of $\mathbf{K}$. $|v\rangle_j$ and $\langle w|_j$ are the right and left eigenvectors of $\mathbf{K}$, respectively, with zero eigenvalue (for a more detailed description of the MAPS method, including the derivation of eqs 3 and 4, see the Supporting Information).

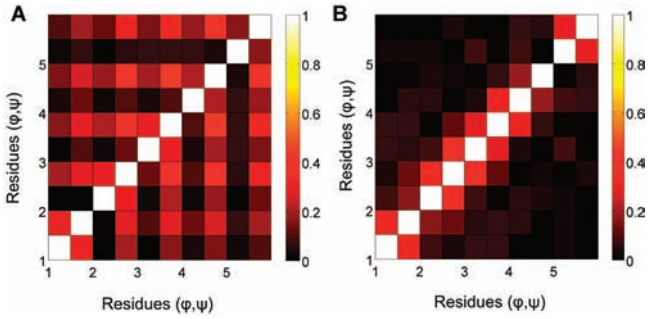


Figure 1. Backbone φ,ψ dihedral angle correlations of alanine pentapeptide, Ala_5, from 2 μs MD simulation. The absolute values of the Pearson correlation coefficients for all pairs of backbone φ,ψ dihedral angles are determined (A) from all conformations of Ala_5 and (B) from the subset of conformations in the all-helical state ($\alpha\alpha\alpha\alpha\alpha$). The all-helical state accounts for 62% of the total population. Along the x (y) axis, the left (bottom) and right (top) fields assigned to each residue correspond to φ and ψ, respectively.

As an independent control for the population shift predicted by eq 3, a 2 μs MD simulation was performed in explicit solvent with residue Ala5 restrained to its coil state (β), followed by the analysis of the change of equilibrium populations. For this purpose, a harmonic restraining potential was applied to both φ and ψ angles of Ala5 of the form:

$$V(\xi) = \frac{1}{2}k_\xi(\xi - \xi_0)^2 \qquad (5)$$

where k_ξ = 328.28 kJ/mol/rad^2, and ξ_0 = −60° or 120° is the equilibrium value for the φ or ψ angles, respectively.

RESULTS AND DISCUSSION

Medium- to Long-Range Correlations of the Peptide Dynamics. In the 2 μs MD trajectory, the Ala_5 peptide effectively samples multiple conformations where each residue can jump between the helical and the coil state, which differs, for example, from the native-basin fluctuation behavior of ubiquitin.[21] However, when focusing on the subensemble where Ala_5 resides in its "native" energy basin of the all-helical state $\alpha\alpha\alpha\alpha\alpha$ accounting for 62% of the total population, the absence of any large medium- and long-range dihedral angle correlations is revealed (Figure 1B). Hence, in its native energy basin, the Ala_5 peptide exhibits, despite its smaller size, dihedral angle correlation characteristics that are qualitatively similar to the ones of ubiquitin.[18,19] By contrast, when the entire MD ensemble of the pentapeptide is analyzed involving conformational jumps dominated by the ψ angle among all 32 different substates, a qualitatively different global correlation pattern emerges (Figure 1A), with $|r|$ = 0.19−0.44 (0.33 on average) between different ψ angles. These types of correlations provide conditions that permit the longer-range transmission of structural signals upon local perturbation.

Global Effect of a Local Conformational Restraint. Any biomolecular process, such as a ligand binding event, can have an effect on the structure and dynamics of a protein. Although such a process often starts out via local interactions with residues in direct contact, it eventually may have a remote or even global effect on the protein involving protein regions that are far beyond the direct contact region with the ligand. In the present study, we model such a local perturbation by a harmonic dihedral potential (eq 5) applied to the φ,ψ angles of the C-terminal residue, Ala5, representing a strong bias of the distribution of the local DOF.

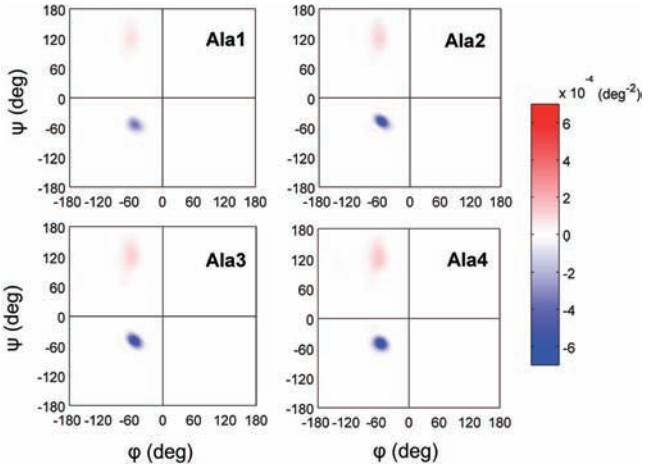


Figure 2. Change of population densities of alanine pentapeptide $\Delta P_{residue}(\varphi,\psi) = P^{(restrained)}_{residue} - P^{(unrestrained)}_{residue}$ in the φ,ψ dihedral space for residues Ala1−Ala4 upon restraining residue Ala5 with the potential of eq 5 during a 2 μs MD trajectory. The changes of the total populations for Ala1, Ala2, Ala3, and Ala4 are 17.5%, 24.6%, 28.1%, and 33.4%, respectively.

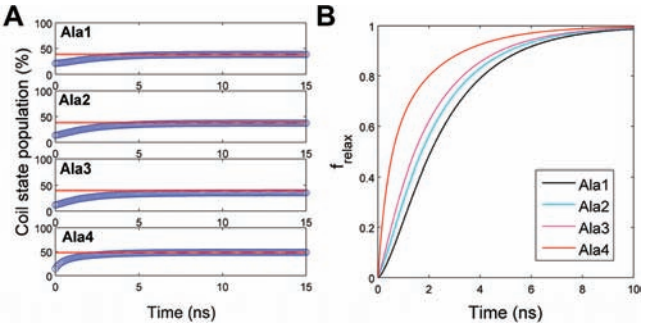


Figure 3. (A) Time-dependence of equilibration after perturbation at time 0 is applied as calculated by the master equation (blue ○). The red dashed lines represent the new populations of individual residues extracted from the restrained simulation. (B) Retardation effects of the structural response as revealed by the time-dependence of populations during relaxation to the new equilibrium for individual residues according to eq 6.

This harmonic potential, which is strictly confined to the φ,ψ angles of residue Ala5, represents an ideally localized perturbation of the system. Interestingly, the effect of such a perturbation is eventually spread over the other four residues Ala1−Ala4, which are unconstrained, as can be seen in Figure 2: the equilibrium populations of individual residues shift toward the coil states, which is apparent even for the N-terminal residue Ala1 (Figure 2). Although this is a sizable effect, it falls short of locking Ala1−Ala4 into their all-β state. The result illuminates at atomic detail the inner working of allostery for this model system where the local structural propensity of an "active site" residue (Ala1) can be substantially changed upon a perturbation at the distant "allosteric site" (Ala5).

Time Course of Equilibration from Master Equation. A benefit of the master equation is that it not only provides equilibrium populations but also the time-dependence of the populations of individual substates as they approach a new equilibrium in response to a perturbation. The accuracy of the

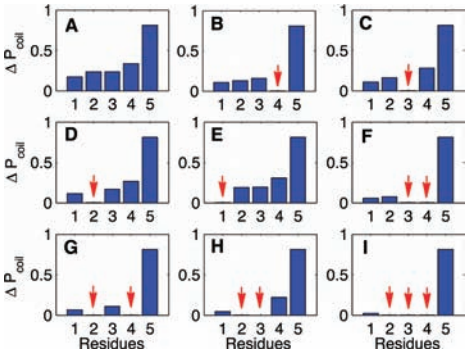


Figure 4. Change of coil state populations ΔP_{coil} of individual residues predicted by the master equation in the presence of dynamic restrictions for selected dihedral angles. Residue Ala5, which is locally restrained to the coil state, shows the largest ΔP_{coil} value. Residues whose conformational transitions are "switched off" are indicated by red arrows.

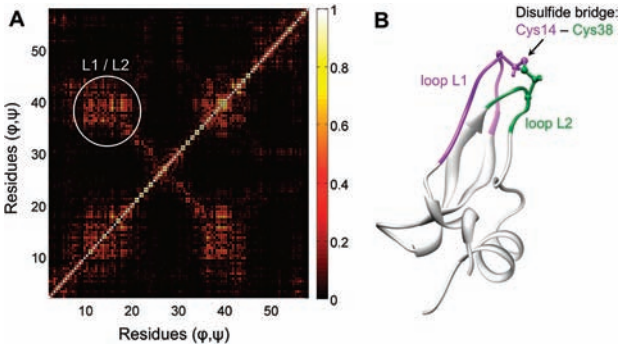


Figure 5. (A) Absolute correlation coefficients $|r|$ of the φ,ψ torsional dynamics of BPTI determined from the 41 250-member molecular ensemble corresponding to the 1.03 ms MD simulation. Correlations between loops L1 and L2 are highlighted by the white circle. (B) The 15 residues belonging to loop L1 (residues 9−18) and loop L2 (residues 36−40) are colored in magenta and green, respectively, whose dynamics is analyzed by the MAPS approach.

predicted population shift can be assessed by comparison with the explicit MD simulation subjected to a perturbation by the application of the restraining potential of eq 5. As shown in Figure 3A, the perturbed system reaches a new equilibrium state on the nanosecond time scale, with the conditional populations derived from the master equation that are in excellent agreement with the ones obtained from the MD simulation.

To analyze retardation effects involving the structural communication between the C-terminal residue, Ala5, and the other residues, the population change of the coil conformation is normalized for each residue, followed by monitoring of the time-dependence of the approach toward equilibrium:

$$f_{relax}(t) = \frac{P_{coil}(t) - P_{coil}(0)}{P_{coil}(t \rightarrow \infty) - P_{coil}(0)} \quad (6)$$

The relaxation process exhibits a multiexponential behavior (Figure 3B) reflecting the complexity of the transition network. Examination of the function f_{relax} in Figure 3B for Ala1−Ala4 reveals the order of effective response times for re-equilibration after perturbation of Ala5: $\tau_{Ala4} < \tau_{Ala3} < \tau_{Ala2} < \tau_{Ala1}$. The perturbation propagates in a diffusive manner rather than in a wave-like manner affecting the residues that are close in terms of primary sequence and spatial distance. The new equilibrium is reached within ~10 ns.

Signal Transmission from C- to N-Terminus. In the polyalanine peptide, residues Ala5 and Ala1 represent the two most distant sites both sequentially and spatially. Therefore, it is instructive to examine the pathways of transmission from Ala5 to Ala1. For this purpose, we selectively switch off the transition of individual residues by setting the corresponding rate constants to zero (see Methods). As shown in Figure 4I, even when residues Ala2−Ala4 are all prohibited from changing their states, residue Ala1 still experiences a small effect of the distant coil bias applied to residue Ala5, indicating a weak, but non-negligible direct long-range coupling between Ala1 and Ala5. When the dynamic restrictions to either one of the residues Ala2−Ala4 are lifted, ΔP_{coil} of Ala1 increases (Figures 4F−H). When the dynamic restrictions to two or more residues of Ala2−Ala4 are released, the increase of ΔP_{coil} of Ala1 continues (Figure 4A−D). In addition, the restriction to any one of the residues Ala1−Ala4 will decrease ΔP_{coil} of the other three residues (Figure 4B−E). This result underpins the distributed nature of the underlying communication network simultaneously involving all parts of the peptide, which is consistent with the multiple pre-existing pathway hypothesis for allosteric regulation.[4]

Application to Millisecond MD Trajectory of BPTI. Because allosteric conformational changes in proteins typically take place on relatively slow time scales,[17,38,39] direct observation of allosteric transitions by standard all-atom MD simulations has traditionally been difficult. However, the recent development of the special-purpose computer Anton[40] has permitted all-atom MD simulations on much longer time scales than previously possible, with which Shaw et al. have performed a simulation of the folded-state dynamics of BPTI over 1 ms.[27] Here, we apply the MAPS approach to this trajectory to explore long-range communication in BPTI.

The trajectory reversibly samples dihedral angle transitions multiple times with their mutual correlations depicted in Figure 5A. A sizable number of nonsequential dihedral angle correlations occur between loop L1 (residues 9−18) and loop L2 (residues 36−40), which are connected by a disulfide bridge between Cys14 and Cys38 (Figure 5B). In experiments, the χ_1 angles of both of these cystines undergo isomerizations on the microsecond−millisecond time scales, thereby enhancing transverse nuclear spin relaxation.[41,42] The sampling of backbone dihedral angle φ,ψ space by the 15 residues on loops L1 and L2 is depicted in Figure 6. They show multimodal distributions preferentially sampling between 2 and 4 distinct regions. For the application of MAPS, the backbone conformations of individual residues are grouped into discrete states, indicated by the cyan circles with 15° radius in Figure 6, where only transitions between the circled regions are counted as a change of state. If each amino acid sampled the conformational regions independently from each other, loops L1 and L2 would occupy 442 368 substates. However, because of interdependencies of the occupancies and limited sampling, only a subset of 2081 states are actually sampled during the 1 ms trajectory.

It is instructive to analyze the mutual dynamic correlations between all of the φ,ψ angles in terms of their pairwise distances. As can be seen in Figure 7, the Pearson correlations tend to diminish at larger distances (>15 Å). Interestingly, a number of moderately large correlations ($r > 0.3$) occur in the medium- to long-distance range of 10−15 Å. These correlations, which are statistically significant as judged from their uncertainties when dividing the trajectory into three consecutive subtrajectories of

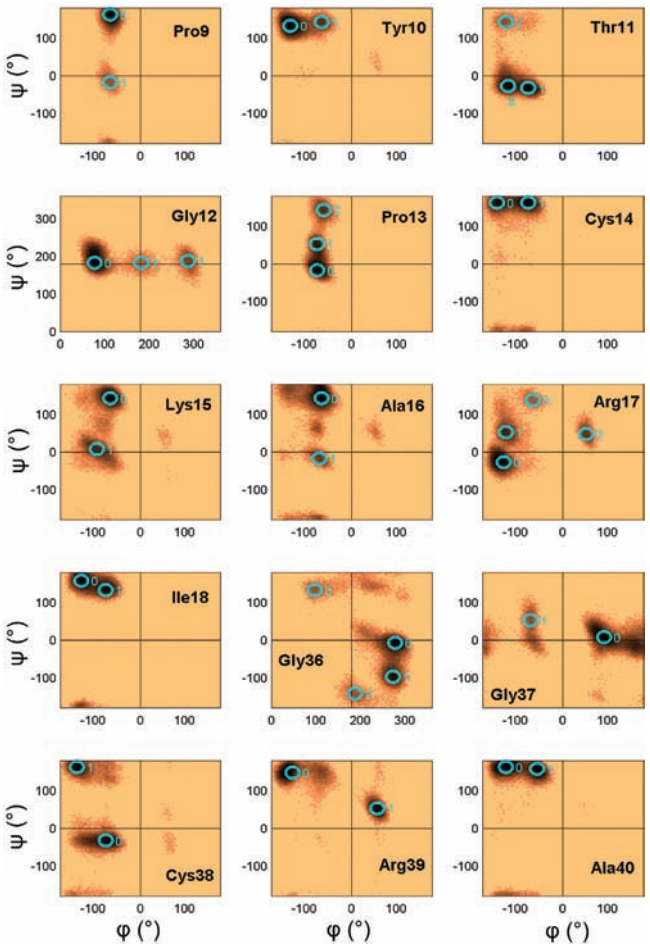


Figure 6. The Ramachandran dihedral angle distributions of the 15 residues of loop L1 (residues 9−18) and loop L2 (residues 36−40). The color shading represents the potential of mean force $-\ln(p)$ per grid element of 5° × 5°, which ranges from $[-\ln(p)]_{min}$ to $+\infty$ for each residue (from black to sandy brown). The backbone conformations of individual residues are grouped into discrete states, indicated by the cyan circles of 15° radius, providing a discrete representation of the system suitable for the MAPS approach. Only transitions between the circled regions are regarded as a change of state.

identical lengths of 0.34 ms each (see Figure S1), involve residue pairs Lys15/Arg39, Arg17/Cys38, and Arg17/Arg39 with $|r| >$ 0.6 located in distant parts (>11 Å) of the two loops.

On the basis of the kinetic matrix **K** constructed from the MD simulation, we perform an in silico experiment, in analogy to the polyalanine peptide, by investigating the allosteric effect of a local perturbation on the side-chain χ_1 dihedral angle of Cys14. As shown in Figure 8A, upon restraining Cys14 χ_1 in the gauche$^-$ conformation, a global population shift of backbone conformations of the two loops is observed whereby the 15 loop residues are affected to a differential degree. As can be seen in Figure 8C, the largest population shift occurs for Cys14 χ_1, which is expected because this is the dihedral angle that was fully constrained to the gauche$^-$ conformation. Remarkably, the backbone conformational states also show large population shifts of up to $\Delta P \approx 50\%$, which belong to residues both in L1 and in L2 (Figure 8A). The χ_1 angle distribution of Cys38, which is the disulfide-bond partner of Cys14, strongly favors the gauche$^+$ over the gauche$^-$ conformation (see Figure S3) at equilibrium, and its

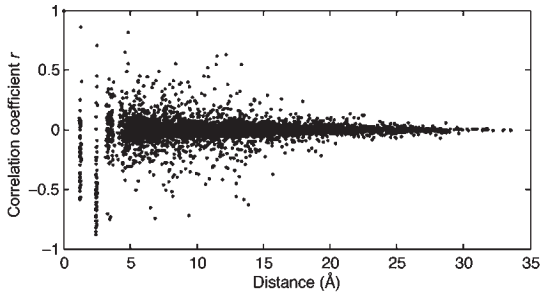


Figure 7. Plot of correlation coefficient r versus distance of all pairs of backbone φ,ψ angles of BPTI. The position of a given dihedral, defined by the atom quartet $atom_1-atom_2-atom_3-atom_4$, is defined as the midpoint between $atom_2$ and $atom_3$. For the calculation of r, the φ,ψ angles are placed in the intervals $\varphi \in [0,360°)$ and $\psi \in [-120°,240°)$. The slow conformational transitions observed during the millisecond BPTI trajectory give rise to several substantial correlations ($|r| > 0.3$) at medium distance range (up to ∼15 Å).

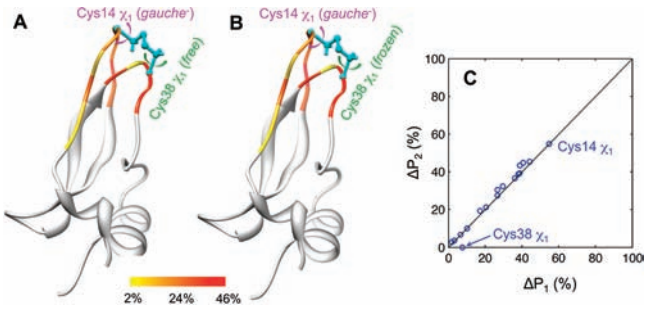


Figure 8. (A,B) The absolute values of population change (ΔP) of individual residues upon a local perturbation, which enforces the gauche$^-$ state of Cys14 χ_1 (A) without and (B) with freezing the transitions of Cys38 χ_1, denoted as ΔP_1 or ΔP_2, respectively. For the residues with more than two states, ΔP refers to the state with the largest population change. (C) Scatter plot of ΔP_2 versus ΔP_1 for the 15 residues (φ,ψ states), Cys14 χ_1, and Cys38 χ_1.

population is shifted by only 8% from gauche$^+$ to gauche$^-$ by the constraint on Cys14 χ_1. This is a manifestation of weak dynamic coupling between these two cystines, consistent with their low Pearson correlation coefficient (r = 0.27). When constraining Cys14 χ_1 and blocking the transitions of Cys38 χ_1, the effect on the population shift pattern is very minor (Figure 8B,C). This implies that the communication between Cys14 χ_1 and dihedral angles of loop L2 can largely bypass the through-disulfide bond pathway.

CONCLUSION

The population shift mechanism has become a preferred view of allostery in biomolecules, but the lack of accurate molecular ensembles has hampered the detailed understanding of allosteric effects in concrete cases. Allosteric propagation of a perturbation via the population shift mechanism can only take place when some of the underlying DOF show a sizable degree of dynamics correlations. Proteins that only display weak correlations over medium- and long-range distances have limited capability for allosteric behavior within the population shift framework. Because the experimental quantification of long-range correlations is very challenging, atomic detail MD simulations have an important role to play.

In this work, we introduce the MAPS formalism as a tool to quantitatively address the origin of allostery. On the basis of time-resolved molecular ensembles generated by long all-atom equilibrium MD simulations, the internal equilibrium dynamics of a polypeptide is translated into a discrete Markov model, which captures the kinetic and thermodynamic equilibrium properties of the unperturbed system at atomic resolution. The concise form of the master equation gives direct access to correlation effects and the systematic identification of the time scales, amplitudes, and pathways of signal transmission. The MAPS formalism is illustrated here for dihedral angles, but it can be applied to other coordinates too. A precondition for the application of MAPS is that conformational substates and their interconversions, which dominate the conditional ensemble after the perturbation, such as ligand binding, must be statistically adequately sampled in the simulation of the free state. With ever longer MD simulations becoming available,[43] this condition is increasingly well fulfilled. In cases where the time scales for allosteric signal transmission are too long to be accessed by current MD, enhanced sampling techniques or a large number of short MD trajectories are valuable options.[23,44] The initial perturbation is modeled in the present work by the application of a local restraining potential to dihedral angles that reside at the allosteric site. Alternatively, one can reweight the free ensemble through the computation of explicit energy changes during the early approach of a ligand to the allosteric site as was demonstrated recently for the Hrs-UIM:ubiquitin complex.[21] This will help elucidate the effects of long-range electrostatic interactions on population shift and allostery.

For the alanine pentapeptide, the master equation predicts how a specific perturbation at the C-terminus propagates throughout the whole molecule, which is validated against an explicit restrained simulation. The result supports a population-shift mechanism of allostery and reveals multiple indirect and direct communication pathways between distant sites. Application of the MAPS approach to a 1 ms trajectory of BPTI exemplifies the medium-range impact of a local perturbation on Cys14 χ_1. Despite the presence of the disulfide linkage (Cys14−Cys38) to loop (L2), the side chain Cys38 χ_1 is not a critical participant in the transmission pathway.

Taken together, allosteric signal propagation within the framework of population shift can take place in systems with a network of sizable torsional correlations as displayed both by the alanine pentapeptide and by the L1/L2 loop region of BPTI. While these correlations indicate soft- to medium-strong coupling behavior between the respective DOF, they can induce significant population changes in remote regions when one or several DOF are constrained. As molecular mechanics force fields, algorithms, and computer hardwares continue to advance, increasingly realistic ensembles of larger biomolecules and their complexes can be constructed and validated against quantitative experiments. Their comprehensive analysis by kinetic models, such as MAPS, is expected to provide important new insights on the many facets of allostery in its role for biomolecular function.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information. Detailed mathematical description of MAPS, additional analysis of BPTI (four figures), and complete ref 40. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
bruschweiler@magnet.fsu.edu

■ ACKNOWLEDGMENT

We thank D. E. Shaw Research for making their MD trajectory of BPTI publicly available. This work was supported by the National Science Foundation (grant MCB-0918362).

■ REFERENCES

(1) Changeux, J. P.; Edelstein, S. J. *Science* **2005**, *308*, 1424–1428.
(2) Hilser, V. J. *Science* **2010**, *327*, 653–654.
(3) Low, C.; Homeyer, N.; Weininger, U.; Sticht, H.; Balbach, J. *ACS Chem. Biol.* **2009**, *4*, 53–63.
(4) de Sol, A.; Tsai, C. J.; Ma, B.; Nussinov, R. *Structure* **2009**, *17*, 1042–1050.
(5) Gunasekaran, K.; Ma, B.; Nussinov, R. *Proteins* **2004**, *57*, 433–443.
(6) Weber, G. *Biochemistry* **1972**, *11*, 864–878.
(7) Stan, G.; Lorimer, G. H.; Thirumalai, D.; Brooks, B. R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 8803–8808.
(8) Selvaratnam, R.; Chowdhury, S.; VanSchouwen, B.; Melacini, G. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 6133–6138.
(9) Zhuravleva, A.; Gierasch, L. M. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 6987–6992.
(10) Monod, J.; Wyman, J.; Changeux, J. P. *J. Mol. Biol.* **1965**, *12*, 88–118.
(11) Koshland, D. E.; Nemethy, G.; Filmer, D. *Biochemistry* **1966**, *5*, 365–368.
(12) Kar, G.; Keskin, O.; Gursoy, A.; Nussinov, R. *Curr. Opin. Pharmacol.* **2010**, *10*, 715–722.
(13) Cui, Q.; Karplus, M. *Protein Sci.* **2008**, *17*, 1295–1307.
(14) Vendruscolo, M. *Nat. Chem. Biol.* **2011**, *7*, 411–412.
(15) Ma, B.; Tsai, C. J.; Haliloglu, T.; Nussinov, R. *Structure* **2011**, *19*, 907–917.
(16) Arora, K.; Brooks, C. L. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18496–18501.
(17) Kern, D.; Zuiderweg, E. R. P. *Curr. Opin. Struct. Biol.* **2003**, *13*, 748–757.
(18) Li, D.-W.; Meng, D.; Brüschweiler, R. *J. Am. Chem. Soc.* **2009**, *131*, 14610–14611.
(19) Fenwick, R. B.; Esteban-Martin, S.; Richter, B.; Lee, D.; Walter, K. F. A.; Milovanovic, D.; Becker, S.; Lakomek, N. A.; Griesinger, C.; Salvatella, X. *J. Am. Chem. Soc.* **2011**, *133*, 10336–10339.
(20) Brüschweiler, R. *Nat. Chem.* **2011**, *3*, 665–666.
(21) Long, D.; Brüschweiler, R. *PLoS Comput. Biol.* **2011**, *7*, e1002035.
(22) Bowman, G. R.; Huang, X.; Pande, V. S. *Cell Res.* **2010**, *20*, 622–630.
(23) Buchete, N.-V.; Hummer, G. *Phys. Rev. E* **2008**, *77*, 030902.
(24) Buchete, N.-V.; Hummer, G. *J. Phys. Chem. B* **2008**, *112*, 6057–6069.
(25) De Sancho, D.; Best, R. B. *J. Am. Chem. Soc.* **2011**, *133*, 6809–6816.
(26) Deng, N. J.; Zheng, W.; Gallicchio, E.; Levy, R. M. *J. Am. Chem. Soc.* **2011**, *133*, 9387–9394.
(27) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341–346.
(28) Nymeyer, H.; Garcia, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13934–13939.
(29) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472–2493.

(30) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(31) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
(32) Garcia, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2782–2787.
(33) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
(34) Hess, B.; Becker, S.; Berendsen, H. J.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
(35) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–725.
(36) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950–1958.
(37) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665–9678.
(38) Hammes, G. G. *Biochemistry* **2002**, *41*, 8221–8228.
(39) Brüschweiler, S.; Schanda, P.; Kloiber, K.; Brutscher, B.; Kontaxis, G.; Konrat, R.; Tollinger, M. *J. Am. Chem. Soc.* **2009**, *131*, 3063–3068.
(40) Shaw, D. E.; et al. *Proceedings of the Conference on High Performance Computing (SC09)*; ACM Press: New York, 2009.
(41) Otting, G.; Liepinsh, E.; Wüthrich, K. *Biochemistry* **1993**, *32*, 3571–3582.
(42) Grey, M. J.; Wang, C.; Palmer, A. G. *J. Am. Chem. Soc.* **2003**, *125*, 14324–14335.
(43) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. *Curr. Opin. Struct. Biol.* **2009**, *19*, 120–127.
(44) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Dill, K. A. *Multiscale Model. Simul.* **2006**, *5*, 1214–1226.