Entropic mechanism of large fluctuation in allosteric transition

Kazuhito Itoh[a,1] **and Masaki Sasai**[a,b]

[a]Department of Applied Physics, Nagoya University, Nagoya, 464-8603, Japan; and [b]Korea Institute for Advanced Study, Seoul 130-722, Korea

Edited by Peter G Wolynes, University of California, San Diego, La Jolla, CA, and approved March 16, 2010 (received for review November 10, 2009)

A statistical mechanical model of allosteric transitions in proteins is developed by extending the structure-based model of protein folding to cases of multiple native conformations. The partition function is calculated exactly within the model and the free-energy surface reflecting allostery is derived. This approach is applied to an example protein, the receiver domain of the bacterial enhancer-binding protein NtrC. The model predicts the large entropy associated with a combinatorial number of preexisting transition routes. This large entropy lowers the free-energy barrier of the allosteric transition, which explains the large structural fluctuation observed in the NMR data of NtrC. The global allosteric transformation of NtrC is explained by the shift of preexisting distribution of conformations upon phosphorylation, but the local structural adjustment around the phosphorylation site is explained by the complementary induced-fit mechanism. Structural disordering accompanied by fluctuating interactions specific to two allosteric conformations underlies a large number of routes of allosteric transition.

free-energy landscape | population shift | statistical mechanical model

Allosteric change of protein conformation is essential for many biological regulatory processes. As possible mechanisms of allostery, two limiting cases have been compared (1–6): One is the population-shift mechanism whereby a protein takes on multiple native conformations prior to the binding of the effector molecule to the protein. The binding of the effector then stabilizes one of the preexisting metastable conformations and makes that conformation the most populated. The other is the induced-fit mechanism whereby the effector binds to a preferential conformation of the protein and then the protein is modulated to form a conformation that was not observed before the effector binding.

Recent experimental data have shown the existence of fluctuations among multiple native conformations prior to the effector binding in several example allosteric proteins such as NtrC (nitrogen regulatory protein C) (7, 8), CheY (9–11), calmodulin (12–14), and adenylate kinase (15–17), which has supported the view that the population-shift mechanism is not exceptional but may be ubiquitous in allosteric transitions (18–23). There still remain, however, important questions to be elucidated: First, the actual mechanism in each allosteric protein may reside between two limits of population shift and induced fit, so that the relative importance of two mechanisms should be analyzed in each case (24). Second, when the population shift dominates the allostery of proteins examined, we should ask how the preexisting fluctuations among multiple native conformations are realized. In this paper, we develop a statistical mechanical model of allostery to analyze these two questions.

To answer these questions for NtrC, for example, several theoretical attempts have been made by using targeted molecular dynamics (25, 26), normal mode analysis (27), discrete path sampling (28), self-guided Langevin dynamics (29), and structure-based and structure-prediction methods (30). However, only the restricted space around certain selected conformations or paths were sampled with the methods of refs. 25–29, and hence those results should be compared with the results of other methods that can sample the more extensive conformational space.

Coarse-grained structure-based models are suitable for such global sampling, and hence two-basin models, the structure-based models that take account of the two low-energy reference conformations, have been developed and applied to allosteric transitions (30–37). Whitford et al. (36), for example, used structure-based potentials that stabilize the apo structure uniquely and then added a perturbation representing the effects of ligand binding to lower the eneregy of the holo structure too. To examine the relative importance of the population-shift and induced-fit mechanisms, however, both apo and holo structures should be treated on equal footing. Such a balanced treatment was considered by Okazaki et al. (24, 32): The interaction potentials in their model are self-consistently switched from those stabilizing one structure to those stabilizing the other. This switching, however, is made uniformly through the protein, so that the calculation beyond such mean-field-like approaches is necessary to describe the spatially nonuniform pathway of allosteric change. It is, therefore, strongly desired to develop a model that can take account of two reference structures on equal footing and describe the nonuniform change in interactions.

In this paper we develop a coarse-grained model that meets these requirements by extending the statistical mechanical model of protein folding first proposed by Wako and Saito (38–45). By extending Bruscolini and Pelizzola's transfer-matrix method (42) and using the method in two steps, we can obtain the exact partition function of this extended model. With the obtained partition function, free-energy landscapes and other physical quantities are readily calculated over the global conformational space, which allows the detailed and transparent analyses of allosteric transitions.

Results

Statistical Mechanical Model of Allostery. We consider a protein that undergoes allosteric transition between two native conformations, "active" (A) and "inactive" (I), and assume that the most stable conformation is changed from I to A by binding the effector. We use I and A as references to describe other conformations: When the backbone and the side-chain structures of the kth residue in the conformation examined are close to those in the A(I) conformation, we write $m_k = \mathrm{A(I)}$, and otherwise, $m_k = \mathrm{D}$. When both the backbone and side-chain structures of the kth residue are close to common structures of the A and I conformations, that residue is classified into type "common" (C). For the residue of type C, we can write either of $m_k = \mathrm{A}$ or $m_k = \mathrm{I}$. We use this redundant expression for C to write down the partition function in a compact form, as will be shown below in Eq. **4**.

The present model is an extension of the model of folding in which the many-body effects are taken into account by considering

Author contributions: K.I. and M.S. designed research; K.I. performed research; K.I. contributed new reagents/analytic tools; K.I. and M.S. analyzed data; and K.I. and M.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: kazuhito@tbp.cse.nagoya-u.ac.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.0912978107/-/DCSupplemental.

the contiguous segment of native-like configurations (38, 39, 41). The present extension succeeds this feature by focusing on segments or "stretches" of configurations. We introduce two projection functions at the kth site, $p_k^{\mathrm{A}}(m_k)$ and $p_k^{\mathrm{I}}(m_k)$. $p_k^{\mathrm{A}}(m_k)$ is the projection onto the A state as $p_k^{\mathrm{A}}(\mathrm{A})=1$ and $p_k^{\mathrm{A}}(\mathrm{I})=p_k^{\mathrm{A}}(\mathrm{D})=0$, and $p_k^{\mathrm{I}}(m_k)$ is the projection onto the I state as $p_k^{\mathrm{I}}(\mathrm{I})=1$ and $p_k^{\mathrm{I}}(\mathrm{A})=p_k^{\mathrm{I}}(\mathrm{D})=0$. We write $p_k^{\mathrm{N}}(m_k)=p_k^{\mathrm{A}}(m_k)+p_k^{\mathrm{I}}(m_k)$ and $p_k^{\mathrm{D}}(m_k)=1-p_k^{\mathrm{N}}(m_k)$. A segment of $\prod_{k=i}^{j}p_k^{\mathrm{N}}(m_k)=1$ is called "N-stretch (i, j)," a segment of $\prod_{k=i}^{j}p_k^{\mathrm{D}}(m_k)=1$ is called "D-stretch (i, j)," and a segment of $\prod_{k=i}^{j}p_k^{\mathrm{A(I)}}(m_k)=1$ is called "A(I)-stretch (i, j)." When the segment is composed of only the type C residues, we regard the segment as either an A- or I-stretch. In this way, the protein conformation is hierarchically described by a mosaic of D- and N-stretches and a mosaic of A- and I-stretches in each N-stretch as shown in Fig. 1. We use this hierarchy to calculate the partition function in two steps: At the first step, the statistical weight of each N-stretch is calculated by solving a two-valued problem of A and I in each N-stretch by borrowing the technique of a two-valued problem of folding (42), and at the second step, the partition function of the entire protein chain is calculated by solving the two-valued problem of N and D.

We refer to a contact between residues in A or I conformation as a native contact, and effects of nonnative contacts are ignored. We write $\Delta_{i,j}^{\mathrm{A(I)}}=1$ when the native contact is formed between ith and jth residues of the A(I) conformation, and $\Delta_{i,j}^{\mathrm{A(I)}}=0$, otherwise. Hamiltonian of the effector-nonbinding (binding) state, $\mathscr{H}_{\alpha=0(1)}$, is

$$\mathscr{H}_{\alpha}(\mathbf{m})=\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}[\mathscr{H}_{i,j}^{0}+\mathscr{H}_{i,j}^{\mathrm{A}}+\mathscr{H}_{i,j}^{\mathrm{I}}]+\mathscr{V}_{\alpha}, \qquad [1]$$

$$\mathscr{H}_{i,j}^{0}=\varepsilon_{i,j}\Delta_{i,j}^{0}\prod_{k=i}^{j}p_k^{\mathrm{N}}(m_k), \qquad [2]$$

$$\mathscr{H}_{i,j}^{\mathrm{A(I)}}=\varepsilon_{i,j}\tilde{\Delta}_{i,j}^{\mathrm{A(I)}}\prod_{k=i}^{j}p_k^{\mathrm{A(I)}}(m_k), \qquad [3]$$

where $\mathbf{m}=(m_1,m_2,\cdots,m_N)$, $\varepsilon_{i,j}$ is the contact energy between ith and jth residues, and N is the total number of residues. We can see from Eqs. **1**–**3** that two conformations A and I are treated on equal footing in Hamiltonian. $\mathscr{H}_{i,j}^{0}$ is the energy gain due to a contact that is common to A and I, $\Delta_{i,j}^{0}\equiv\Delta_{i,j}^{\mathrm{A}}\Delta_{i,j}^{\mathrm{I}}=1$. This energy is gained only when all the residues from i to j belong to the same N-stretch. $\mathscr{H}_{i,j}^{\mathrm{A(I)}}$ is the energy gain due to a contact that is not common to A and I with $\tilde{\Delta}_{i,j}^{\mathrm{A(I)}}\equiv\Delta_{i,j}^{\mathrm{A(I)}}(1-\Delta_{i,j}^{0})=1$, which arises when all the residues from i to j belong to the same A(I)-stretch. $\mathscr{V}_1-\mathscr{V}_0=\mathscr{V}$ is the energy change around the binding site induced by the effector binding, and we put $\mathscr{V}_0=0$ without loss of generality. Then, $\mathscr{V}_1\neq 0$ modulates the statistical weight of stretches, leading to the global conformational transition from I to A. We note here that $\Delta_{i,j}^{\mathrm{A(I)}}$ and m_k are defined by the atomistic configuration of both side-chain and backbone structures, so that the difference in the side-chain packing brings about difference in contacts, $q_{ij}^{0}=\Delta_{i,j}^{0}\prod_{k=i}^{j}p_k^{\mathrm{N}}(m_k)$ or $\tilde{q}_{ij}^{\mathrm{A(I)}}=\tilde{\Delta}_{i,j}^{\mathrm{A(I)}}\prod_{k=i}^{j}p_k^{\mathrm{A(I)}}(m_k)$. Such difference in q_{ij}^{0} or $\tilde{q}_{ij}^{\mathrm{A(I)}}$ can mediate changes in interactions from the effector-binding site to the area remote from it to induce the large conformational change there, even in the case that there is no significant backbone conformational change in between that area and the effector-binding site.

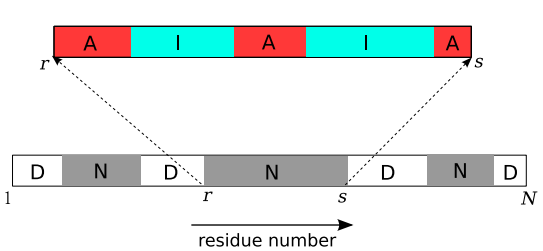

Fig. 1. An example of protein conformation represented with D, A, and I-stretches.

In the present model, the conformational transition is characterized by the trade-off among the energy of attractive interactions in N-stretches, entropy arising from the larger number of nonnative configurations in D-stretches, and entropy from the combinatorial number of arrangements of A-, I-, and D-stretches. Such balance between energy and entropy can be described by the partition function,

$$Z_{\alpha}=\sum_{\{m_k\}}\prod_{k=1}^{N}(d_k)^{p_k^{\mathrm{N}}(m_k)}\prod_{k=1}^{N}(g_k)^{p_k^{\mathrm{D}}(m_k)}\exp[-\beta\mathscr{H}_{\alpha}(\mathbf{m})], \qquad [4]$$

which is a sum over all the configurations of $\{m_k|m_k=\mathrm{A},\mathrm{I},\mathrm{D}\}$ with $\beta=1/k_{\mathrm{B}}T$. Here, g_k is the number of nonnative configurations of the kth residue, and d_k is a factor to evaluate the weight of the structure of type C. We use $d_k=1/2$ when the kth residue takes the structure of type C and $d_k=1$ for otherwise. It is natural to assume that contacts in the segment of type C satisfy $\tilde{\Delta}_{i,j}^{\mathrm{A(I)}}=0$, so that only the term $\mathscr{H}_{i,j}^{0}$ should contribute. Hence, the segment of type C gives rise to the same energy gain irrespective of whether it is regarded as an A- or I-stretch. In calculating Eq. **4**, the segment of type C is counted twice as A- and I-stretches, respectively, but this overcounting is canceled by the factor $d_k=1/2$, so that physically the residue is interpreted to have a unique structure. In this way, we have a consistent description of conformational states with the three-letter representation of $m_k=$ A, I, and D. With this representation, the protein conformation is described in a hierarchical way as in Fig. 1, which allows the use of the transfer-matrix method in a hierarchically decomposed way. Further explanation of the partition function is given in *SI Text* and Fig. S1.

NtrC, an Example Allosteric Protein. We apply this method to the allosteric transition in NtrC as an example (7, 8, 46). The receiver domain NtrC is a single-domain α/β protein with 124 residues, and its I conformation (PDB ID: 1dc7) is switched to the A conformation (PDB ID: 1dc8) by phosphorylation (46) (Fig. 2). Recent NMR study by Volkman (7) et al. and Gardino et al. (8) presented the clear evidence for the population shift between I and A conformations: They found a strong correlation between the phosphorylation-driven activation of NtrC and the backbone dynamics of micro- to milli-second timescale in an area remote from the phosphorylation site in the dephosphorylated state.

We define the native contact as $\Delta_{k,l}^{\mathrm{A(I)}}=1$ when a heavy atom other than hydrogen in the kth residue and that in the lth residue with $l>k+2$ are closer than 5Å in the A(I) conformation. For simplicity, we assume that contact energies are $\varepsilon_{i,j}=\varepsilon\leq 0$ for residues distant from the phosphorylation site Asp54. We also assume $\varepsilon_{i,j}=\varepsilon$ for $\Delta_{i,j}^{0}=1$ irrespective of whether the pair $[i,j]$ is far from or near Asp54. The change of the atomistic environment around Asp54 is described by change in other Asp54-neighbor interactions. We write a set of Asp54-neighbor pairs as $\mathscr{P}$ and define the pair $[i,j]$ to belong to $\mathscr{P}$ when i or j is Asp54 or when i or j is the residue that contacts Asp54. See SI Text and Table S1 for more on the native contacts and Asp54-neighbor residues. For those Asp54-neighbor pairs, $\varepsilon_{i,j}=0$ for the pair of $\tilde{\Delta}_{i,j}^{\mathrm{A}}=1$ and $\varepsilon_{i,j}=\varepsilon$ for $\tilde{\Delta}_{i,j}^{\mathrm{I}}=1$ in the dephosphorylated state, and $\varepsilon_{i,j}=\gamma\varepsilon$ for the pair of $\tilde{\Delta}_{i,j}^{\mathrm{A}}=1$ and $\varepsilon_{i,j}=0$ for $\tilde{\Delta}_{i,j}^{\mathrm{I}}=1$ in the phosphorylated state. We thus have $\mathscr{V}=\sum_{[i,j]\in\mathscr{P}}[\gamma\varepsilon\tilde{\Delta}_{i,j}^{\mathrm{A}}\prod_{k=i}^{j}p_k^{\mathrm{A}}(m_k)-\varepsilon\tilde{\Delta}_{i,j}^{\mathrm{I}}\prod_{k=i}^{j}p_k^{\mathrm{I}}(m_k)]$. For simplicity, we adopt an approximate expression of

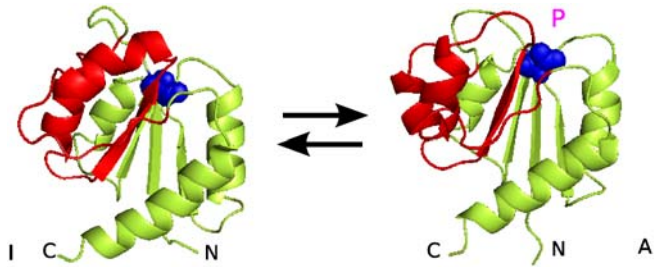

Fig. 2. Conformations of NtrC. NtrC undergoes the allosteric transition between inactive (dephosphorylated, I) and active (phosphorylated, A) conformations. Red colored are residues in "3445 face" (the surface region composed of helices and strands, α3, β4, α4, and β5) (46) and the blue colored is Asp54.

$g_k/d_k = 2g$ when the kth residue has a similar native backbone configuration in A and I, and $g_k/d_k = g - 1$ when it has different A and I backbone configurations. The definition of the similarity of the backbone structure is explained in *SI Text*. When the PDB data for A and I conformations are available, $\Delta_{k,l}^{A(I)}$ and the residues to be $g_k/d_k = 2g$ are determined. Then, the remaining parameters of the model are γ, ε/k_BT, and g.

Population Shift vs. Induced Fit in NtrC. In Fig. 3 the free-energy surfaces $F_\alpha(x,n)$ are shown in the two-dimensional space of n and x for the dephosphorylated ($\alpha = 0$) and phosphorylated ($\alpha = 1$) cases. Here, n is the order parameter of the folding transition defined by the ratio of the number of residues that take the A or I structure to N. $x = M_A/M$ is the order parameter of the allosteric transition with M_A being the number of residues that take the backbone structure specific to the A configuration and M being the total number of residues that can take the backbone structure specific to the A configuration. $(x,n) = (0,1)$ for the I conformation and (1,1) for the A conformation. The free-energy minimum at $(x,n) \approx (0.2, 0.4)$ in Fig. 3 *A* and *B* is the unfolded state. There are free-energy basins corresponding to folding intermediates at $(x,n) \approx (0.3, 0.75)$ in Fig. 3*A* and at $(x,n) \approx (0.3, 0.65)$ in Fig. 3*B*. Although these intermediates can kinetically trap the folding trajectories, Fig. 3 *A* and *B* and Fig. S2 *A–D* show

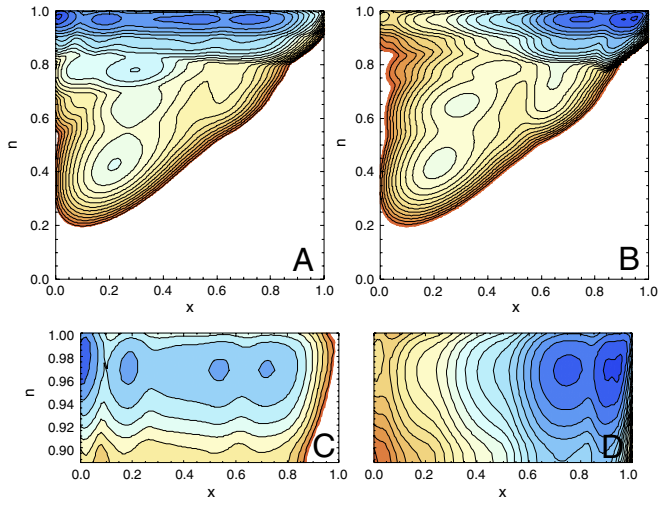

Fig. 3. Free energy of the allosteric transition of NtrC. The calculated free-energy surfaces in the dephosphorylated (*A*) and phosphorylated (*B*) states are shown in the two-dimensional space of reaction coordinates n and x, where n is the order parameter of folding and x is the variable representing the degree of formation of the active structure. Closeups of regions of $n \approx 1$ of the free energy are shown in the dephosphorylated (*C*) and phosphorylated (*D*) states. Contour is drawn in every $2k_BT$. $\varepsilon/k_BT = -0.8$, $\gamma = 2.0$, and $g = 4.5$.

that the folding transition is thermodynamically a two-state transition from the unfolded state to the I or A conformation.

In Fig. 3 *A* and *B*, we can find the valley of low free energy at $n \approx 1$. This feature shows that the allosteric transition observed by the variation of x can take place without significant decrease of n. In other words, the unfolding fluctuation monitored by n is not significant in the allosteric transition, but as will be explained later in *Discussion*, the partial disordering defined by the decrease of contacts plays an important role in the allosteric transition of NtrC.

Closeups of regions of $n \approx 1$ of Fig. 3 *A* and *B* are given in Fig. 3 *C* and *D*, respectively. In Fig. 3*C*, in addition to a free-energy minimum corresponding to I, there are three local minima at $(x,n) \approx (0.2, 0.97)$, (0.55,0.97), and (0.75,0.97) corresponding to intermediates of the allosteric transition, which are mutually connected by low free-energy barriers, forming a large basin of intermediates. The conformation of NtrC in the dephosphorylated state can, therefore, thermally fluctuate within this basin. As shown in Fig. 3 *C* and *D*, the shallow minimum at $(x,n) \approx (0.75, 0.97)$ in the dephosphorylated state becomes a distinct minimum upon phosphorylation, and the population of conformations residing at this minimum substantially increases due to this development of the minimum in the phosphorylated state. This result supports the interpretation of the NMR data for the population-shift mechanism in NtrC (7).

In Fig. 3*D*, the lowest free-energy minimum is at the A conformation of $(x,n) \approx (0.95, 0.97)$. This minimum does not exist before phosphorylation as shown in Fig. 3*C*. The population of conformations at this minimum is small in the dephosphorylated state, so that the last step of allosteric transition, transition from the intermediate at $(x,n) \approx (0.75, 0.97)$ to the A conformation should be regarded as an induced-fit process. In this intermediate at $x \approx 0.75$, residues around Asp54 locally take the I structure, so that this last step corresponds to the local structural adjustment at around Asp54. In this way, the present calculation shows that the population-shift mechanism dominates the allosteric transition of NtrC as suggested by experiment (7), but the local adjustment complements the population shift at the last step. This last step is an induced transition to the state newly created by phosphorylation, so that we may call this step the complementary induced fit.

Dependence of the results on parameters γ, ε/k_BT, and g is explained below. The complementary induced fit is sensitive to the change in the value of γ that represents the extent to which the phosphorylation reaction proceeds. The A conformation of $x \approx 1$ becomes lowest in free energy for $\gamma > \gamma_c$, but the intermediate at $x \approx 0.75$ is lowest for $\gamma < \gamma_c$. For $-0.95 \le \varepsilon/k_BT \le -0.65$, γ_c can be approximately fitted by $\gamma_c \approx 6.0\varepsilon/k_BT + 6.5$. Increase in γ from $\gamma < \gamma_c$ to $\gamma > \gamma_c$ thus induces the transition of the local configuration adjustment around the phosphorylated site.

As shown in Fig. S2, a decrease in $|\varepsilon|/k_BT$ destabilizes both I and A conformations and stabilizes the unfolded state. With a smaller decrease in $|\varepsilon|/k_BT$, however, intermediates of the allosteric transition at $n \approx 1$ are also stabilized. Stabilization of intermediates in dephosphorylated NtrC can be shown by the ratio of the equilibrium population of intermediates to that in the basin of I, which is calculated at $n = 1$ as $K \equiv \sum_{x>x^\ddagger} Z_0(x,1) / \sum_{x<x^\ddagger} Z_0(x,1)$, where $x^\ddagger \approx 0.1$ is the transition state between I and one of intermediates. As shown in Fig. 4*A*, $10^{-3} < K < 1$ for $-0.9 \le \varepsilon/k_BT \le -0.75$, for which I is stable. In this condition the barrier height of excitation from I to intermediates is about 5.3 to $10.3k_BT$ as shown in Fig. 4*B*. By assuming the Arrhenius law with the prefactor $k_0 \approx 10^8$ s^{-1} (41), the rate of the conformational exchange between I and intermediates is estimated to be 5.0×10^3 to 6.7×10^5 s^{-1}, and hence the upper limit of the rate is $\sim 10^6$ s^{-1}, which is consistent with the observed NMR relaxation in micro- to milli-second timescale (7).

As shown in Fig. S3, a change of g induces the similar effect to that caused by the change of $|\varepsilon|/k_BT$. With the fixed ε/k_BT, the

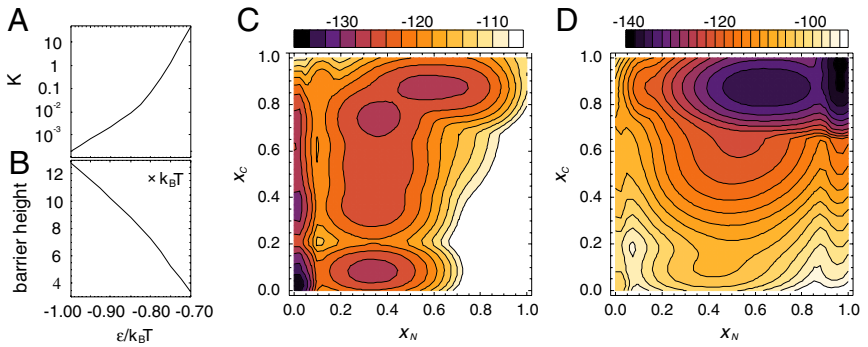

Fig. 4. Population ratio, barrier height, and free-energy surfaces of NtrC. (*A*) The equilibrium population ratio and (*B*) the barrier height between inactive and intermediate states as functions of ε/k_BT. Two-dimensional free-energy surfaces in (*C*) dephosphorylated and (*D*) phosphorylated states are shown with coordinates x_N and x_C, which represent the formation of the active structure in the N and C-terminal halves of NtrC. $M_N = 41$ and $M_C = 46$. Contour is drawn in every $2k_BT$. $\varepsilon/k_BT = -0.8$ and $\gamma = 2.0$. n is fixed to be 1.

global minimum is at $n \approx 1$ for $g < g_c$, whereas it is at the unfolded state for $g > g_c$. When $\varepsilon/k_BT \leq -0.75$, $g_c > 5.25$. The entropic factor for taking nonnative structures should be $\ln g \approx 1.5$ to properly describe the free-energy surface of folding/unfolding (47), so that we use $g = 4.5$ in the following, but the results of the transition between I and A are insensitive to g as far as $g < g_c$.

Large Conformational Fluctuations. As shown in Fig. 3 *C* and *D*, the free-energy landscape has a distinct valley of $n \approx 0.97$ both in the dephosphorylated and phosphorylated states, which allows large conformational fluctuations along this valley. These fluctuations can be further analyzed by imposing the approximate constraint of $n = 1$. Indeed, 3–4 residues at the N- and C-terminal edges, which have no contacts with other residues, take denatured structures in this valley, and hence a good approximation for the partition function is obtained by $Z_\alpha \approx \text{const.} \times Z_\alpha(n = 1)$ to describe the conformational fluctuation between I and A.

The NMR data showed the presence of large conformational exchange at the C-terminal half before phosphorylation (preexisting fluctuation) and at the N-terminal half after phosphorylation (7). To compare the calculated results with these experimental data, we show in Fig. 4 *C* and *D* free-energy surfaces of $n = 1$ with the two-dimensional coordinates of x_N and x_C, where $x_{N(C)} = M_{AN(C)}/M_{N(C)}$ with $M_{AN(C)}$ being the number of residues whose backbone structure takes the A configuration in the N(C) terminal half and $M_{N(C)}$ being the number of residues that have different backbone structures in the N(C)-terminal half of A and I: $(x_N, x_C) = (0, 0)$ for the I conformation and (1,1) for the A conformation.

In the dephosphorylated state the free-energy surface has a valley in the region $x_N \approx 0$ and $0 < x_C < 0.8$. By surmounting the barrier from I, fluctuation along this valley is induced, through which I and A structures are exchanged at the C-terminal half. In addition to this valley, the free energy of the dephosphorylated NtrC has a shallow basin at around $0.2 < x_N < 0.8$ and $0.2 < x_C < 0.8$, which lowers the free-energy barrier of the allosteric transition and allows large fluctuations in the dephosphorylated state. The complex free-energy surface with multiple basins of the dephosphorylated NtrC shows that the allosteric transition from I to A states has multiple pathways. In the phosphorylated state, the free-energy surface has a basin at around $0.5 < x_N < 0.8$ and $0.75 < x_C < 0.95$ and shows that the N-terminal half of NtrC fluctuates more than the C-terminal half, as has been observed in the NMR data (7).

As demonstrated in the Φ-value analyses of protein folding (48), structural features along the transition pathway can be analyzed by the response to the local modulation of interaction. When $\varepsilon_{i,j}$ of the contact with $\tilde{\Delta}^A_{i,j} = 1$ is modulated to $\varepsilon_{i,j} + \delta\varepsilon$ with $|\delta\varepsilon|/k_BT \ll 1$, for example, the free energy constrained at $\mathbf{x} = x$ or $\mathbf{x} = (x_N, x_C)$ should change from $F_\alpha(\mathbf{x}, 1)$ to $F_\alpha(\mathbf{x}, 1) + \Delta F^A_{\alpha ij}(\mathbf{x})$. This response of $F_\alpha(\mathbf{x}, 1)$ is quantified by $\xi^A_{\alpha ij}(\mathbf{x}) \equiv \Delta F^A_{\alpha ij}(\mathbf{x})/\delta\varepsilon$, which represents the statistical average of the contact formation $\tilde{q}^A_{ij}$ and satisfies $0 \leq \xi^A_{\alpha ij}(\mathbf{x}) \leq 1$ by definition. The average of $\xi^A_{\alpha ij}(\mathbf{x})$ at the jth site, $0 \leq \xi^A_{\alpha j}(\mathbf{x}) \leq 1$, is the order parameter of how well the A structure is developed at the jth site with a given $\mathbf{x}$. All neighbors of the jth residue are in the same A stretch with the jth residue when $\xi^A_{\alpha j}(\mathbf{x}) = 1$. In a similar way, $\xi^I_{\alpha ij}(\mathbf{x})$ is defined for contacts with $\tilde{\Delta}^I_{i,j} = 1$, and the order parameter of the I structure development, $\xi^I_{\alpha j}(\mathbf{x})$, can be defined.

By focusing on the preexisting fluctuations, the order parameter $\xi^A_{\alpha j}(x)$ along the one-dimensional coordinate x in the dephosphorylated state ($\alpha = 0$) is shown in Fig. 5. We can see that as x is increased, the A structure is formed from the C-terminal half as suggested by the free-energy landscape of Fig. 4*C*. The A structure first develops in residues 91–95 (α4) at $x \approx 0.2$, and corresponding to the shallow basin in Fig. 4*C*, the A structure develops at $x \approx 0.55$ in residues 89–101 (α4 and β5) and at $x \approx 0.75$ in residues 61–101 (α3, β4, α4, and β5) that constitute the surface region called "3445 face" of NtrC (46). Compared in Fig. 5 is the experimental data of R_{ex} taken from (7). The experimental data shows the large preexisting conformational exchange at around Asp54 and at the 3445 face, which is consistent with the calculated data of Fig. 5.

Further refined structural information is obtained from $\xi^{A(I)}_{\alpha ij}(x)$, which can be interpreted as the order parameter of how well the contact specific to the A(I) conformation develops. As shown in Fig. S4*A*, major differences between A and I conformations are denser distributions of contacts at around Asp54 in the I conformation and at the 3445 face in the A conformation. Shown in Fig. S4 *B–D* are $\xi^A_{\alpha ij}(x)$ and $\xi^I_{\alpha ij}(x)$ in the dephosphorylated state ($\alpha = 0$) at various x. At small x, contacts of the A conformation in the 3445 face begin to develop but the Asp54-neighbor contacts around $i = 15$, 54, 62, and 63 remain as those in the I conformation. As x increases, the

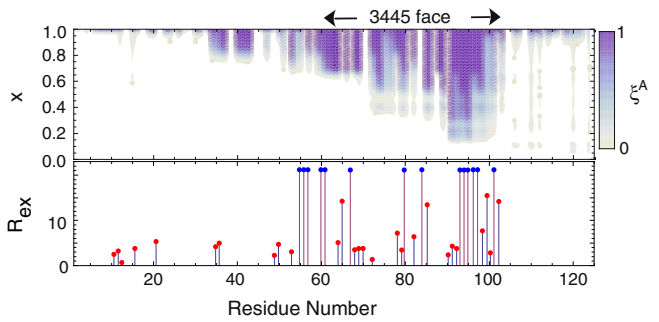

Fig. 5. Structural features of preexisting fluctuation in NtrC. (Top) The order parameter of formation of the active structure at each residue in the dephosphorylated state $\xi^A_{0j}(x)$ is represented as a function of x. n is fixed to be 1. Even in conformations near to the inactive conformation ($x \approx 0$), the active structure appears as fluctuation around the 3445 face, which corresponds to the region where the large values of R_{ex} have been observed with NMR. (Bottom) Red dots are the values of R_{ex} observed with NMR (7). For the data of blue dots, R_{ex} has been larger than a threshold in the experimental data but their precise values have not been determined (7).

Asp54-neighbor contacts in the I-conformation are lost and the structure around Asp54 gradually turns into that of the A-conformation. The Asp54-neighbor contacts in the A-conformation around $i = 9$–12, however, scarcely develop even at $x \approx 0.75$ and are formed only at the last step to approach the A conformation.

The structural change in the dephosphorylated state is shown with $\xi^{A}_{aj}(x_N, x_C)$ and $\xi^{I}_{aj}(x_N, x_C)$ in Fig. S5. In the phosphorylated state, on the other hand, the approach to the A conformation is shown on a three-dimensional conformation in Fig. 6. Starting from the state near the I conformation, phosphorylation triggers the destabilization of the I structure-like feature around Asp54 and in the C-terminal region (Fig. 6*A*), and as the transition proceeds along the free-energy slope of Fig. 4*D*, the A structure begins to develop in the C-terminal region (Fig. 6*B*) and spreads over the 3445 face (Fig. 6*C*). The A structure develops around Asp54 and in the N-terminal region at the last step to approach the A conformation. See Fig. S6 for further details. These detailed features of conformation change could be quantitatively tested further by mutagenesis studies as in the Φ-value analyses of folding (48).

Entropic Mechanism of Preexisting Fluctuation. A remaining important question is the mechanism of how such large preexisting fluctuation is realized. To analyze the mechanism, we decompose free energy $F_\alpha(\mathbf{x})$ at $n = 1$ and $\mathbf{x} = (x_N, x_C)$ into energy $E_\alpha(\mathbf{x}) = \langle \mathscr{H}_\alpha \rangle^\alpha_{\mathbf{x}}$, where $\langle \cdots \rangle^\alpha_{\mathbf{x}}$ is the average taken with $\mathscr{H}_\alpha$ under the constraint of $\mathbf{x}$ in the phosphoform α (see *SI Text*), and entropy $S_\alpha(\mathbf{x})$ as $F_\alpha(\mathbf{x}) = E_\alpha(\mathbf{x}) - TS_\alpha(\mathbf{x})$. $S_\alpha(\mathbf{x})$ can be further decomposed into the energy-dependent part $S_{e\alpha}(\mathbf{x})$ and the energy-independent part as

$$S_\alpha(\mathbf{x}) = S_{e\alpha}(\mathbf{x}) + k_B \ln[\Omega_0 \Omega_b(\mathbf{x})], \quad [5]$$

$$S_{e\alpha}(\mathbf{x}) = k_B \ln\{\langle \exp[\beta(\mathscr{H}_\alpha - E_\alpha(\mathbf{x}))]\rangle^\alpha_{\mathbf{x}}\}^{-1}, \quad [6]$$

where $\Omega_0 = \prod_{k=1}^{N} d_k$ and $\Omega_b(\mathbf{x}) = \binom{M_N}{M_N x_N}\binom{M_C}{M_C x_C} 2^{N-M_N-M_C}$ and $\Omega_0\Omega_b$ is the combinatorial number of configurations of A and I at $\mathbf{x}$. $S_{e\alpha}(\mathbf{x})$ is negative and represents the entropic cost due to the effective reduction of the number of available contact patterns (47). Energy dependence of $S_{e\alpha}(\mathbf{x})$ reflects cooperativity of forming nonlocal correlation of interactions in $\mathscr{H}_\alpha$. Reduction of the number of possible side-chain configurations due to the constraint of fixed $\mathbf{x}$ is represented in this term.

Energy $E_{\alpha=0}(\mathbf{x})$ and entropy $S_{\alpha=0}(\mathbf{x})$ of the dephosphorylated state are shown in Fig. 7 *A* and *B*. In the shallow basin of intermediates shown in Fig. 4*C* at around $0.2 < x_N < 0.8$ and $0.2 < x_C < 0.8$, $E_{\alpha=0}(\mathbf{x})$ significantly increases due to the loss of contacts of Δ^{I}_{ij} with insufficient formation of $\tilde{\Delta}^{A}_{ij}$ as shown in Fig. S4. The targeted molecular dynamics simulation suggested that the transiently formed nonnative interactions play roles to compensate this increase in energy (8, 26). Though nonnative interactions should affect the transition kinetics, further careful examination is necessary to see whether they match the large increase in energy of Fig. 7*A*. In the present model, this increase in energy is compensated by increase in entropy $S_{\alpha=0}(\mathbf{x})$ as shown in Fig. 7*B*, leading to the small free-energy barrier. This large entropy comes from a combinatorially large number of mosaic configurations of A and I stretches along the protein chain. In other words, there are combinatorially many possible routes from I to A conformations that lower the free-energy barrier and induce the large preexisting conformational fluctuation. This mechanism of entropic lowering is intrinsically nonlinear associated with the backbone and side-chain conformational fluctuations ranging over the entire protein. This mechanism in NtrC is different from the cracking mechanism proposed by Miyashita et al. (49) in which the free-energy lowering is due to the entropic gain in nonnative backbone configurations that are different from either A or I structure: In our expression the entropic gain associated with cracking should be $k_B \ln g_k$ in D-stretches, whereas the entropic gain proposed here is represented by Eq. **5**. We should note that the free-energy valley of intermediates in the region $x_N \approx 0$ and $0 < x_C < 0.8$ is energetically stabilized due to the balance of loss of $\tilde{\Delta}^{I}_{ij}$ and acquisition of $\tilde{\Delta}^{A}_{ij}$ along this valley. As shown in Fig. 7*C*, $S_{e\alpha} = 0(\mathbf{x})$ is less negative in the region of $x_C \approx 0$, 0.6, and 0.9, which gives rise to the free-energy basins of $x_C \approx 0.1$, 0.7, and 0.9 in Fig. 4*C*, so that $S_{e\alpha} = 0(\mathbf{x})$ explains the detailed mechanism of the preexisting conformational fluctuations.

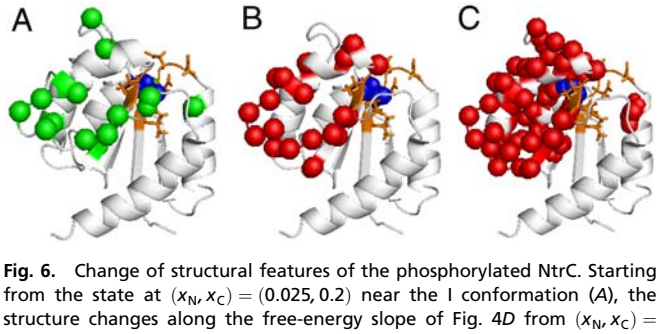

Fig. 6. Change of structural features of the phosphorylated NtrC. Starting from the state at $(x_N, x_C) = (0.025, 0.2)$ near the I conformation (*A*), the structure changes along the free-energy slope of Fig. 4*D* from $(x_N, x_C) = (0.025, 0.78)$ (*B*) to $(x_N, x_C) = (0.68, 0.87)$ (*C*). In *A*, green represents residues at which the I structure-like feature is decreased upon phosphorylation as $\xi^{I}_{1j}(x_N, x_C) - \xi^{I}_{0j}(x_N, x_C) < -0.2$. Red colored residues having the A structure-like feature with $\xi^{A}_{1j}(x_N, x_C) > 0.2$ and $\xi^{I}_{1j}(x_N, x_C) < 0.2$ appear in *B* and spread over the 3445 face in *C*. Blue and orange colored are Asp54 and Asp54-neighbor residues, respectively.

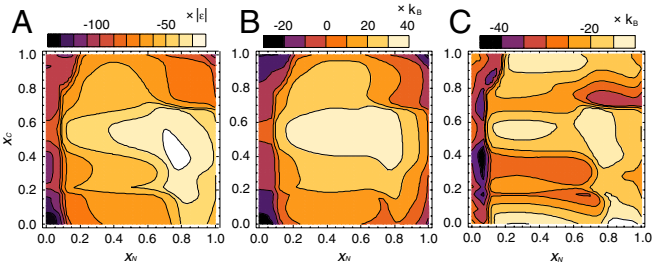

Fig. 7. Free energy of the dephosphorylated NtrC in Fig. 4*C* is decomposed into energy and entropy. (*A*) Energy, (*B*) entropy, and (*C*) the Hamiltonian dependent part of entropy $S_{e\alpha} = 0$ are represented in the two-dimensional space of x_N and x_C. In *B* the constant term $k_B \ln \Omega_0$ is omitted from the values exhibited.

Discussion

An interesting problem is the possibility of partial unfolding during the allosteric transition (49). As shown in *Results*, the allosteric transition and preexisting fluctuations take place without significant decrease of n. When analyzing breakage and formation of contacts, however, we can see the importance of disordered configurations that are not expressed by the change in n. Shown in Fig. S7 *A* and *B* are free-energy surfaces represented on the two-dimensional space of Q_A and Q_I, where $Q_{A(I)}$ is the number of contacts formed between pairs with $\tilde{\Delta}^{A(I)}_{ij} = 1$. Because n is fixed to be $n = 1$ in Fig. S7, the number of contacts formed between pairs with $\Delta^{0}_{ij} = 1$ is kept its maximal value. Especially in the dephosphorylated state of Fig. S7*A*, we can see that there are broad areas of low free energy in between A and I conformations, so that Q_A and Q_I fluctuate largely in the dephosphorylated state, which is consistent with the results that there are a combinatorially large number of routes. With a similar two-dimensional representation of free-energy surface, existence of multiple routes was reported also with the full-atom simulation for a small protein (50). With fluctuations of Q_A and Q_I, contacts are broken and formed multiple times. Breakage and formation

of contacts and associated disordering and ordering in side-chain packing are necessary condition to allow the large number of transition routes. Such disordering may correspond to the "partial unfolding" reported in simulations of NtrC (28, 29) and resembles the simulated fluctuations in protein kinase A (37).

Also interesting is the result showing that intermediates of the allosteric transition at $n \approx 1$ shown in Fig. 3 can be the on-pathway intermediates of the folding/unfolding transition. In these intermediates, the native contacts with $\tilde{\Delta}_{ij}^{A(I)} = 1$ are partially lost but the contacts with $\Delta_{ij}^{0} = 1$ are maintained. This sharing of the intermediates in allosteric and folding/unfolding transitions suggests that information about the allosteric transition can be obtained from unfolding experiments of adding denaturant or changing pH. The statistical mechanical model proposed here can be applied to allosteric transitions that do not exhibit the backbone conformational changes but show changes in the side-chain packing (22, 23). Such an allosteric transition can be analyzed by using, for example, Q_A and Q_I, which are defined by both the side-chain and backbone configurations.

In summary the statistical mechanical model reproduced the observed fluctuations in the exemplified allosteric transition and suggested the intrinsically nonlinear mechanism of entropic lowering of the transition barrier. The population of conformations shifts through the large fluctuations over this barrier, but the complementary induced-fit also works to locally adjust the conformation. We showed the importance of a combinatorially large number of preexisting routes, which underlie the extensively large entropy of the preexisting fluctuation. By projecting these many routes on two-dimensional reaction coordinates as in Fig. 4*C*, we still found a variety of preexisting pathways connecting multiple basins of intermediates as their importance has been suggested in (21). The proposed method opens wide applications for various types of allosteric transition of single- and multidomain proteins (22, 23, http://protein3d.ncifcrf.gov/tsai/allostery/, http://graylab.jhu.edu/allostery/), and further kinetic analyses should be possible by extending the local equilibrium approach of Zamparo et al. (51) for the present model.

ACKNOWLEDGMENTS. The authors thank Prof. D. Wemmer for critical reading of the manuscript. This work was supported by Grants-in-aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology, Japan, and from the Japan Society for the Promotion of Science.

1. Goodey NM, Benkovic SJ (2008) Allosteric regulation and catalysis emerge via a common route. *Nat Chem Biol* 4:474–482.
2. Swain JF, Gierasch LM (2006) The changing landscape of protein allostery. *Curr Opin Struct Biol* 16:102–108.
3. Ma B, Kumar S, Tsai CJ, Nussinov R (1999) Folding funnels and binding mechanisms. *Protein Eng* 12:713–720.
4. Tsai CJ, Kumar S, Ma B, Nussinov R (1999) Folding funnels, binding funnels, and protein function. *Protein Sci* 8:1181–1190.
5. Tsai CJ, Ma B, Nussinov R (1999) Folding and binding cascades: Shifts in energy landscapes. *Proc Natl Acad Sci USA* 96:9970–9972.
6. Boehr DD, Nussinov R, Wright PE (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* 5:789–796.
7. Volkman BF, Lipson D, Wemmer DE, Kern D (2001) Two-state allosteric behavior in a single-domain signaling protein. *Science* 291:2429–2433.
8. Gardino AK, et al. (2009) Transient non-native hydrogen bonds promote activation of a signaling protein. *Cell* 139:1109–1118.
9. Barak R, Eisenbach M (1992) Correlation between phosphorylation of the chemotaxis protein CheY and its activity at the flagellar motor. *Biochemistry* 31:1821–1826.
10. Moy FJ, et al. (1994) Assignments, secondary structure, global fold, and dynamics of chemotaxis Y protein using three- and four-dimensional heteronuclear (13C, 15N) NMR spectroscopy. *Biochemistry* 33:10731–10742.
11. Simonovic M, Volz K (2001) A distinct meta-active conformation in the 1.1-Å resolution structure of wild-type apoCheY. *J Biol Chem* 276:28637–28640.
12. Barbato G, Ikura M, Kay LE, Pastor RW, Bax A (1992) Backbone dynamics of calmodulin studied by N-15 relaxation using inverse detected 2-dimensional NMR-spectroscopy-the central helix is flexible. *Biochemistry* 31:5269–5278.
13. Johnson CK (2006) Calmodulin, conformational states, and calcium signaling. A single-molecule perspective. *Biochemistry* 45:14233–14246.
14. Torok K, Tzortzopoulos A, Grabarek Z, Best SL, Thorogate R (2001) Dual effect of ATP in the activation mechanism of brain Ca^{2+}/calmodulin-dependent protein kinase II by Ca^{2+}/calmodulin. *Biochemistry* 40:14878–14890.
15. Hanson JA, et al. (2007) Illuminating the mechanistic roles of enzyme conformational dynamics. *Proc Natl Acad Sci USA* 104:18055–18060.
16. Henzler-Wildman KA, et al. (2007) Intrinsic motions along an enzymatic reaction trajectory. *Nature* 450:838–844.
17. Henzler-Wildman KA, et al. (2007) A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* 450:913–916.
18. Kern D, Zuiderweg ER (2003) The role of dynamics in allosteric regulation. *Curr Opin Struct Biol* 13:748–757.
19. Gunasekaran K, Ma BY, Nussinov R (2004) Is allostery an intrinsic property of all dynamic proteins?. *Proteins* 57:433–443.
20. Henzler-Wildman K, Kern D (2007) Dynamic personalities of proteins. *Nature* 450:964–972.
21. del Sol A, Tsai CJ, Ma B, Nussinov R (2009) The origin of allosteric functional modulation: Multiple pre-existing pathways. *Structure* 17:1042–1050.
22. Daily MD, Gray JJ (2007) Local motions in a benchmark of allosteric proteins. *Proteins* 67:385–399.
23. Tsai CJ, del Sol A, Nussinov R (2008) Allostery: Absence of a change in shape does not imply that allostery is not at play. *J Mol Biol* 378:1–11.
24. Okazaki K, Takada S (2008) Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms. *Proc Natl Acad Sci USA* 103:11844–11849.
25. Hu X, Wang Y (2006) Molecular dynamic simulations of the N-terminal receiver domain of NtrC reveal intrinsic conformational flexibility in the inactive state. *J Biomol Struct Dyn* 23:509–518.
26. Lei M, et al. (2009) Segmented transition pathway of the signaling protein nitrogen regulatory protein C. *J Mol Biol* 392:823–836.
27. Liu MS, et al. (2008) Coarse-grained dynamics of the receiver domain of NtrC: Fluctuations, correlations and implications for allosteric cooperativity. *Proteins* 73:218–227.
28. Khalil M, Wales DJ (2008) Pathways for conformational change in nitrogen regulatory protein C from discrete path sampling. *J Phys Chem B* 112:2456–2465.
29. Damjanović A, Garciá-Moreno BE, Brooks BR (2009) Self-guided Langevin dynamics study of regulatory interactions in NtrC. *Proteins* 76:1007–1019.
30. Lätzer J, Shen T, Wolynes PG (2008) Conformational Switching upon phosphorylation: A predictive framework based on energy landscape principles. *Biochemistry* 47:2110–2122.
31. Best RB, Chen YG, Hummer G (2005) Slow protein conformational dynamics from multiple experimental structures: The helix/sheet transition of arc repressor. *Structure* 13:1755–1763.
32. Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG (2006) Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc Natl Acad Sci USA* 105:11182–11187.
33. Hyeon C, Lorimer GH, Thirumalai D (2006) Dynamics of allosteric transitions in GroEL. *Proc Natl Acad Sci USA* 103:18939–18944.
34. Takagi F, Kikuchi M (2007) Structural change and nucleotide dissociation of myosin motor domain: Dual go model simulation. *Biophys J* 93:3820–3827.
35. Lu Q, Wang J (2008) Single molecule conformational dynamics of adenylate kinase: Energy landscape, structural correlations, and transition state ensembles. *J Am Chem Soc* 130:4772–4783.
36. Whitford PC, Gosavi S, Onuchic JN (2008) Conformational transitions in adenylate kinase allosteric communication reduces misligation. *J Biol Chem* 283:2042–2048.
37. Hyeon C, Jennings PA, Adams JA, Onuchic JN (2009) Ligand-induced global transitions in the catalytic domain of protein kinase A. *Proc Natl Acad Sci USA* 106:3023–3028.
38. Wako H, Saito N (1978) Statistical mechanical theory of protein conformation I. general considerations and application to homopolymers. *J Phys Soc Jpn* 44:1931–1938.
39. Wako H, Saito N (1978) Statistical mechanical theory of protein conformation II. folding pathway for protein. *J Phys Soc Jpn* 44:1939–1945.
40. Go N, Abe H (1981) Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. Formulation. *Biopolymers* 20:991–1011.
41. Muñoz V, Eaton WA (1999) A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc Natl Acad Sci USA* 96:11311–11316.
42. Bruscolini P, Pelizzola A (2002) Exact solution of the Muñoz-Eaton model for protein folding. *Phys Rev Lett* 88:258101.
43. Itoh K, Sasai M (2004) Dynamical transition and proteinquake in photoactive yellow protein. *Proc Natl Acad Sci USA* 101:14736–14741.
44. Itoh K, Sasai M (2006) Flexibly varying folding mechanism of a nearly symmetrical protein: B domain of protein A. *Proc Natl Acad Sci USA* 103:7298–72303.
45. Itoh K, Sasai M (2008) Cooperativity, connectivity, and folding pathways of multidomain proteins. *Proc Natl Acad Sci USA* 105:13865–13870.
46. Kern D, et al. (1999) Structure of a transiently phosphorylated switch in bacterial signal transduction. *Nature* 402:894–898.
47. Itoh K, Sasai M (2009) Multidimensional theory of protein folding. *J Chem Phys* 130:145104.
48. Fersht A (1999) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (Freeman & Company, New York).
49. Miyashita O, Onuchic JN, Wolynes PG (2003) Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc Natl Acad Sci USA* 100:12570–12575.
50. Wu S, Zhuravlev PI, Papoian GA (2008) High resolution approach to the native state ensemble kinetics and thermodynamics. *Biophys J* 95:5524–5532.
51. Zamparo M, Pelizzola A (2006) Kinetics of the Wako-Saito-Muñoz-Eaton model of protein folding. *Phys Rev Lett* 97:068106.