THE PROTEIN SOCIETY

Disordered linkers in multidomain allosteric proteins: Entropic effect to favor the open state or enhanced local concentration to favor the closed state?

Maodong Li,[1] Huaiqing Cao,[2] Luhua Lai,[1,2,3] and Zhirong Liu[1,2,3]*

[1]Center for Quantitative Biology, Peking University, Beijing, 100871, China
[2]College of Chemistry and Molecular Engineering, Peking University, Beijing, 100871, China
[3]State Key Laboratory for Structural Chemistry of Unstable and Stable Species, Beijing National Laboratory for Molecular Sciences (BNLMS), Peking University, Beijing, 100871, China

Received 28 March 2018; Accepted 24 June 2018
DOI: 10.1002/pro.3475
Published online 00 Month 2018 proteinscience.org

Abstract: There are many multidomain allosteric proteins where an allosteric signal at the allosteric domain modifies the activity of the functional domain. Intrinsically disordered regions (linkers) are widely involved in this kind of regulation process, but the essential role they play therein is not well understood. Here, we investigated the effect of linkers in stabilizing the open or the closed states of multidomain proteins using combined thermodynamic deduction and coarse-grained molecular dynamics simulations. We revealed that the influence of linker can be fully characterized by an effective local concentration $[B]_0$. When K_d is smaller than $[B]_0$, the closed state would be favored; while the open state would be preferred when K_d is larger than $[B]_0$. We used four protein systems with markedly different domain–domain binding affinity and structural order/disorder as model systems to understand the relationship between $[B]_0$ and the linker length as well as its flexibility. The linker length is the main practical determinant of $[B]_0$. $[B]_0$ of a flexible linker with 40–60 residues was determined to be in a narrow range of 0.2–0.6 m*M*, while a too short or too long length would dramatically decrease $[B]_0$. With the revealed $[B]_0$ range, the introduction of a flexible linker makes the regulation of weakly interacting partners possible.

Keywords: multidomain protein; local concentration; disordered linkers; coarse-grained molecular dynamics simulation

Short statement

In this study, coarse-grained molecular dynamics simulations were combined with thermodynamic deduction to reveal the role of disordered linker in multidomain allosteric proteins. It was shown that the influence of linker can be fully characterized by an effective local concentration, which was determined to be as high as 0.2–0.6 m*M* in four markedly different protein systems, making the regulation of weakly interacting partners possible.

Abbreviations: IDRs, intrinsically disordered regions.

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Ministry of Science and Technology of the People's Republic of China; Grant number: 2015CB910300; Grant sponsor: National Natural Science Foundation of China; Grant number: 21633001.

**Correspondence to:* Zhirong Liu, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China. Email: liuzhirong@pku.edu.cn

Introduction

Protein allostery is common in biological processes, where a signal at "another site" modifies the activity/function of a protein. It was originally defined as cooperation between different chains, and its underlying molecular mechanisms in oligomeric proteins are well known as the Monod–Wyman–Changeux (MWC) model[1] and the Koshland–Nemethy–Filmer (KNF) model[2] in terms of a transition between two conformations. Later, it was recognized that conformational

changes in a monomeric protein also result in allosteric regulation.[3–5] Actually, allosteric communication not only occurs via structural transition, but also comes from changes in protein dynamics.[6,7] Theoretically speaking, all nonfibrous proteins are potentially allosteric when the conformational redistribution is not due to direct binding to the functional sites.[8–10] In this sense, it is not surprising that intrinsically disordered regions (IDRs), which lack ordered structures but still possess important biological functions in cell signaling and gene transcription,[11,12] were found to be widely involved in allosteric regulation.[13–16] Some IDRs show disorder-to-order transitions when they bind with their targets, and the allosteric mechanism can be described with the ensemble allostery model (EAM).[17] However, if IDRs are not induced to transform into ordered structures, can they regulate the allosteric process and what is the underlying mechanism?

Actually, IDRs exist in many multidomain allosteric proteins without a disorder-to-order transition. For example, Polo-like kinase1 (PLK1), a serine/threonine kinase controlling cell signaling through phosphorylation, contains a long disordered linker of 51 residues between the kinase domain and the polo-box domain.[18–20] Similarly, PIN1,[21,22] CaN,[23] PKA,[24] PCK,[25] SRC,[26], and CDK2[27] are auto-inhibited allosteric kinases with different length of inter-domain linkers. Generally, the basic architecture of most multidomain allosteric proteins is composed of two domains connected by a disordered linker, and the two domains can adopt a closed (bound) or open (unbound) state, where the equilibrium is affected by the allosteric ligand. The closed-to-open transition enables the protein function, which has been described as a "Bi-stable Switch."[10] A general analysis of the stability difference between the open and closed states, and the influence of ligand binding is plotted in Figure 1. The central quantity is the free energy difference between the open and closed states written as

$$\Delta G_{\mathrm{Closed-Open}} = -RT \ln \frac{P_{\mathrm{Open}}}{P_{\mathrm{Closed}}}, \qquad (1)$$

where, R is the gas constant, T is the temperature, and P_{Closed} and P_{Open} are the percentage of the closed and open states. $\Delta G_{\mathrm{Closed-Open}} > 0$ for the auto-inhibited kinases (left panels in Fig. 1), where the closed state is more stable than the open state without the allosteric ligand (or covalent modifications), and both the substrate and allosteric ligand bind to the open state, but not to the closed state (e.g., the binding pockets are buried in the closed state at the interdomain interface). The binding of an allosteric ligand shifts the equilibrium to the open state, and thus favors substrate binding (activation), resulting in the observed allosteric effect. The disordered linker plays an essential role in ensuring an adequate range of stability for the open/closed state (Supporting Information Fig. S1). Recently, Tompa has proposed a novel entropic model to explain the role of IDRs in allosteric effects, where it was emphasized that a disordered linker contributes an entropic effect to the free energy distribution of a multidomain system, preferring an open state.[13] On the other hand, there exists an opposite case (right panels in Fig. 1), where the open state is more stable without an allosteric ligand, and the allosteric ligand binds to the closed state to shift the open/closed equilibrium. For the latter case, the disorder linker is often thought to increase the local concentration of different domains and enhance the binding affinity of domains, preferring a closed state.[28–30] Here it seems confusing whether the added disordered linker prefers the open or closed state. The allosteric effect occurred only in the left and right panels in Figure 1, but not in the weak-binding pathway.

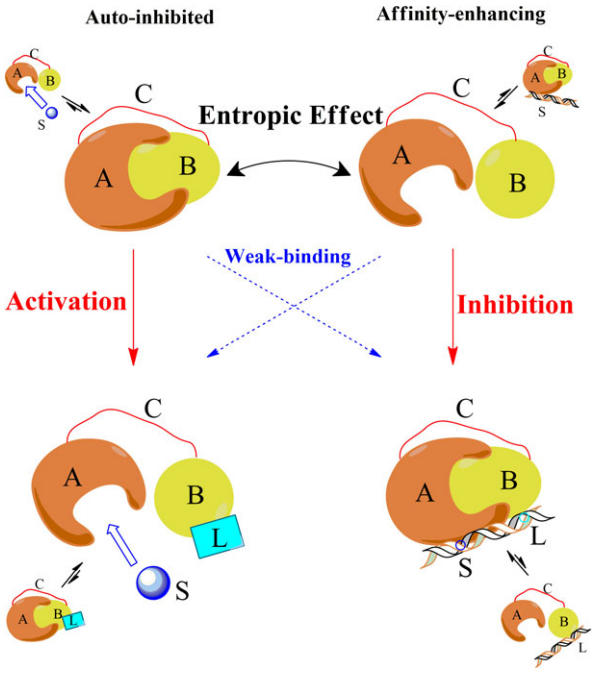

Figure 1. Schematic diagram illustrating the allosteric effects in multidomain proteins with a disordered linker. A protein is composed of two domains (A, the functional domain, and B, the allosteric domain) with mutual attraction, connected by a disordered linker C. The top panels represent the cases without an allosteric ligand, while the bottom panels represent the cases in which an allosteric ligand (L) is added to bind domain B. The left panels represent auto-inhibited systems such as kinase-like proteins, where the ensemble without a ligand is dominated by the closed state, so that its active site in domain A is inaccessible to the substrate S, and the addition of allosteric ligand L changes it to the open state. The right panels represent affinity-enhancing systems such as the DNA-binding proteins, where the ensemble without a ligand is dominated by the open state. The entropic effect of linker C influences the equilibrium between the open and closed states. Adding L leads to four possible pathways (indicated by red and blue arrows) from the initial ensemble to the final ensemble. The allosteric process only exists in the red pathways.

In this article, we combined coarse-grained molecular dynamics simulations and thermodynamic deduction to clarify the role of a disordered linker in $\Delta G_{\text{Closed–Open}}$ of multidomain allosteric proteins. We will show that the role of linker can be fully measured by a characteristic concentration $[B]_0$, which interplays with the domain–domain binding affinity K_d to enable different allosteric pathways: when $[B]_0 > K_d$, the closed state is preferred with $\Delta G_{\text{Closed–Open}} > 0$ where a mechanism described in the left panels in Figure 1 is applicable; when $[B]_0 < K_d$, the open state is preferred with $\Delta G_{\text{Closed–Open}} < 0$ to enable the pathway in the right panels in Figure 1.

Methods

Coarse-grained simulation

Coarse-grained models have been widely used in protein studies,[31–35] especially for large conformation movements.[36–38] We adopt a coarse-grained native-centric Gō-like model,[39–42] to calculate the free energy profile of a protein with two domains connected with a disordered linker. Non-bonded interactions were defined with different adjustable strengths for intradomain interaction, interdomain interaction,[39,43,44] and an extra nonspecific interaction within a disordered chain.[45] As shown in our previous studies, a disordered protein chain can be treated as an isotactic polymer in an equivalent field-effect solution both in the theoretical model [46] and molecular dynamics simulations.[37] The potential energy of the continuum C_α coordinates of the amino acid residues is written as:

$$\begin{aligned}
V_{\text{total}} &= V_{\text{streching}} + V_{\text{bending}} + V_{\text{torsion}} + V_{\text{non-bonded}}^{\text{intra-domain}} \\
&+ V_{\text{non-bonded}}^{\text{inter-domain}} + V_{\text{non-bonded}}^{\text{with-linker}} = \sum_{i,\text{bonds}} K_r\left(r_i - r_i^{(0)}\right)^2 \\
&+ \alpha\sum\nolimits_{i,\text{angles}} K_\theta\left(\theta_i - \theta_i^{(0)}\right)^2 \\
&+ \alpha\sum\nolimits_{i,\text{dihedrals}} \left\{K_{\varphi,1}\left[1-\cos\left(\varphi_i - \varphi_i^{(0)}\right)\right]\right. \\
&\left. + K_{\varphi,3}\left[1-\cos 3\left(\varphi_i - \varphi_i^{(0)}\right)\right]\right\} \\
&+ \alpha\left\{\sum_{\substack{i<j-3\\ \text{native}}}^{\text{intra-domain}} \varepsilon\left[5\left(\frac{r_{ij}^{(0)}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{ij}^{(0)}}{r_{ij}}\right)^{10}\right]\right. \\
&\left. + \sum_{\substack{i<j-3\\ \text{non-native}}}^{\text{intra-domain}} \varepsilon\left(\frac{r_{\text{rep}}}{r_{ij}}\right)^{12}\right\} \\
&+ \beta\left\{\sum_{\substack{i<j-3\\ \text{native}}}^{\text{inter-domain}} \varepsilon\left[5\left(\frac{r_{ij}^{(0)}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{ij}^{(0)}}{r_{ij}}\right)^{10}\right]\right. \\
&\left. + \sum_{\substack{i<j-3\\ \text{native}}}^{\text{inter-domain}} \varepsilon\left(\frac{r_{\text{rep}}}{r_{ij}}\right)^{12}\right\} \\
&+ \left\{\delta\sum\nolimits_{i<j-3}^{\text{with-linker}} \varepsilon\exp\left[-\frac{\left(r_{ij}-\sigma\right)^2}{2}\right]\right. \\
&\left. + \alpha\sum\nolimits_{i<j-3}^{\text{with-linker}} \varepsilon\left(\frac{r_{\text{wl}}}{r_{ij}}\right)^{12} + \alpha\sum\nolimits_{i<j-3}^{\text{linker-domain}} \varepsilon\left(\frac{r_{\text{rep}}}{r_{ij}}\right)^{12}\right\}
\end{aligned} \tag{2}$$

where, r_i, θ_i, φ_i, and r_{ij} are the virtual bond length, bond angle, torsion angle and non-bonded spatial distance defined by C_α coordinates, respectively; $r_i^{(0)}$, $\theta_i^{(0)}$, $\varphi_i^{(0)}$, and $r_{ij}^{(0)}$ are the corresponding native values taken from the PDB structure. Strength parameters were set as $K_r = 100\varepsilon$, $K_\theta = 20\varepsilon$, $K_{\varphi,1} = \varepsilon$, and $K_{\varphi,3} = 0.5\varepsilon$, with $\varepsilon = 1.0$, as described in previous studies.[39–41] The distance parameters were set as $r_{\text{rep}} = 4.0$ Å, $r_{\text{wl}} = 9.0$ Å, and $\sigma = 11.2$ Å to describe nonspecific interactions (Supporting Information Fig. S2).

Non-bonded parameters are vital for simulating conformation changes. Intra-domain interaction parameter α was set to 1.2 as an implicit-water system as mentioned before.[39] Inter-domain interaction parameter β was fixed in simulations according to the experimental binding constant K_d between domains, or alternatively, simply fixed at $\beta = 1.0$. In either case, a reweighting method (as described below) was used to extract the properties of the system under any other β or K_d values.[39] The remaining with-linker interaction parameter δ was the only variable to control linker flexibility, which is affected by the solvent condition.[45,46] A larger β results in tighter domain binding, whereas larger δ results in more collapsed linker conformations and a shorter persistence length (Supporting Information Fig. S3). The flexibility and chain dimension of the linker can be determined by experiments such as small-angle X-ray scattering (SAXS)[37,47] and smFRET.[48,49]

The closed–open transition equilibrium was simulated in the form of Langevin dynamics with the same integration time-step and identical parameterization of mass, viscosity, and the random force as specified in our previous studies.[36,39]

Accelerating method

The free energy profiles were obtained by the umbrella sampling method which contains bias potentials and histogram techniques.[43] We adopt a modified reaction coordinate Q_w to reflect the free energy information.[39] Q_w for a given conformation is defined as:

$$Q_w(\{r_i\},\{r_j\}) = Q_b^{(N)} Q_b(\{r_i\},\{r_j\}) - w\sum_{i\in A, i\in B}^{NCS}\left(r_{ij} - r_{ij}^{(N)}\right), \tag{3}$$

where, $Q_b^{(N)}$ is the number of inter-domain contacts between domain A and B in the native structure. r_{ij} and $r_{ij}^{(N)}$ are the distances between residues i and j in the given conformation and in the PDB, respectively. The constant w was set to 0.02 Å^{-1}. When Q_w is positive, it approximately equals to the native contact number of the given conformation; otherwise, the absolute value of negative Q_w is approximately

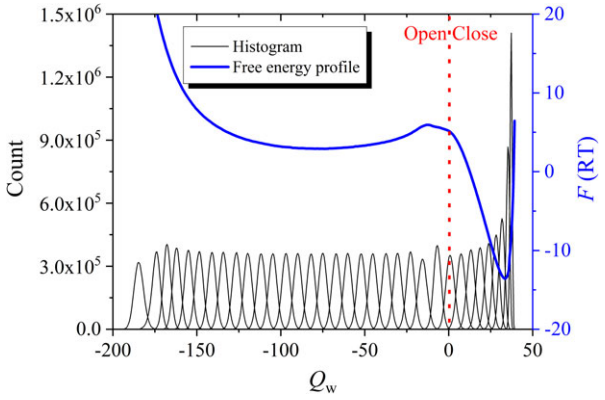

Figure 2. Umbrella sampling method used in this study, taking a linker (N = 60, δ = 0.2) for example: partial histogram results of different bias potentials are shown as black peaks, and the free energy profile is represented in a blue curve as a function of Q_w. $Q_w > 0$ for the closed state, with its value approximately corresponding to the inter-domain contact number; $Q_w < 0$ for the open state, with $-Q_w$ approximately proportional to the domain separation distance. The red dotted line is the boundary between the open and closed states.

proportional to the mass center distance between the two domains.

About 24 parallel trajectories were performed under a harmonic bias potential of:

$$V_{\text{bias}} = a_{\text{bias}}\left(Q_w - Q_{w,\text{bias}}\right)^2, \tag{4}$$

where, a_{bias} was set to 0.02 and $Q_{w,\text{bias}}$ is the harmonic minimum position that varies with different trajectories. For each trajectory, the system is equilibrated in 10^7 steps and the total step is 10^8. The convergence of simulation is fast (see Supporting Information Fig. S4). According to the principle of statistical mechanics, we assembled population distributions to give the overall distribution and free energy of the original system (without the bias potential).[43] A typical free energy profile has two valleys (see Fig. 2), corresponding to the open and the closed states, respectively.

After the free energy profile at one inter-domain interaction strength β value was achieved, the profiles under other β values can be obtained efficiently using a reweighting method based on the principle of statistical mechanics.[39] The population distribution is given as:

$$p(x,U) \propto f(x)\exp\left(-\frac{U}{RT}\right), \tag{5}$$

where, x is any conformational property we want to calculate (namely Q_b and Q_w here) and U is the conformational energy. Therefore, the population distribution under a different interaction strength β' value can be calculated as:

$$p(x,U(\beta')) = p(x,U(\beta))\exp\left(-\frac{U(\beta')-U(\beta)}{RT}\right). \tag{6}$$

Protein systems

In order to differentiate the influence of linker from that of domains, we considered four groups of protein pairs from the literature,[39] namely, Barnase/Barstar (PDB code: 1BRS),[50] BMP/BMP-receptor (PDB code: 2QJ9),[51] CheY/CheA (PDB code: 1A0O),[52] S-protein/S-peptide (PDB code:1D5E),[53] with some of their experimental facts listed in Table I. Two of them are order–order protein pairs, and others are order–disorder protein pairs, considering both strong and weak binding affinity for possible situations. A disordered linker with various lengths and flexibility was added to connect two domains in simulations. For the sake of contrast, the systems without added linker were also simulated, and the simulated K_d was determined as described previously.[39]

Results and Discussion

Determinants of $\Delta G_{\text{closed–open}}$ in multidomain proteins with a disordered linker

The calculated $\Delta G_{\text{Closed–Open}}$ under various conditions for 1BRS is summarized in Figure 3. For usual linkers in a coiled state with δ = 0.2, a linker length of N = 60 rendered the closed state the most stable for the wild-type case (β = 1.35), while the stabilization effect was weakened by shorter or longer linkers. The variation in the range of $\Delta G_{\text{Closed–Open}}$ caused by N changing [indicated by black arrows in Fig. 3(a)] between 30 and 120 residues was about 1.2 RT, which is not sensitive enough to change the pathway of the regulation process in the wild-type case with $\Delta G_{\text{Closed–Open}} \approx 13\ RT$. However, a very short linker (N = 19) caused a significant stability loss in the closed state of about 5.0 RT compared with the most stable case of N = 60 [Fig. 3(c)]. In contrast, the variation in β [indicated by blue arrows in Fig. 3(a)] was more efficient in changing the pathway of the regulation process, which can be implemented by mutation or allosteric ligand binding at the domain–domain interface. If an adding allosteric activating ligand (probably together with interface mutation) is able to decrease the binding affinity by 9.0 RT, the system with a rigid linker would switch from the closed state to the open state for a kinase-like function, as indicated by the blue arrow on the left in Figure 3(a), and a typical pathway-shift was observed.

A disordered linker mainly plays its entropic effect on allosteric regulation when it is expanded, but only contributes to conformation restriction when it is collapsed. When the linker flexibility parameter, δ, (which is related to the nonspecific attraction and can be modified by a denaturant,[46] for instance) increased, a coil-globule transition of the linker occurred at a rough range of $\delta \in$ [0.4, 0.6] [Fig. 3(d)].

Table I. Summary of the Considered Protein Systems[a]

PDB	Domain A (length)	Domain B (length)	Structure	K_d (M)	R_{ee} (Å)	Ref.
1BRS	Barnase (87)	Barstar (108)	Order–order	1.3E-14	48	[50]
2QJ9	BMP (93)	BMP-receptor (103)	Order–disorder	1.0E-8	30	[51]
1A0O	CheY (70)	CheA (128)	Order–order	3.0E-7	49	[52]
1D5E	S-protein (14)	S-peptide (101)	Order–disorder	6.0E-12	15	[53]

[a]The data is for domain–domain complexes before a disordered linker is added to connect two domains. The chain length listed in brackets is measured as the number of residues in the PDB file. R_{ee} is the distance between the C-terminal residue of domain A and the N-terminal residue of domain B in the PDB structure, which sets the constraint for the end-to-end distance of the linker in the closed state.

This lead to a notable decrease in $\Delta G_{\text{Closed–Open}}$ of over 6 RT [Fig. 3(b) and Supporting Information Fig. S5]. Therefore, the linker flexibility is a sensitive parameter in the regulation process [indicated by black arrows in Fig. 3(b)].

Characterization of the linker's influence: the effective local concentration $[B]_0$

The contribution of the linker to $\Delta G_{\text{Closed–Open}}$ can be deduced based on a thermodynamic analysis as follows. We consider the transition equilibriums in both the without-linker protein pair and the with-linker system:

$$\begin{array}{ccc} \text{AB} & \overset{K_d}{\longleftrightarrow} & \text{A}+\text{B} \\ \widehat{\text{AB}} & \overset{\Delta G_{\text{Closed-open}}}{\longleftrightarrow} & \text{A}\sim\text{B} \end{array} \tag{7}$$

Here A, B, AB, A~B, and $\widehat{\text{AB}}$ represents the dissociated protein domain A, the dissociated protein domain B, the protein complex AB, the open state of with-linker system, and the closed state of with-linker system, respectively. The equilibriums are determined by

$$K_d = \frac{[\text{A}][\text{B}]}{[\text{AB}]} \tag{8}$$

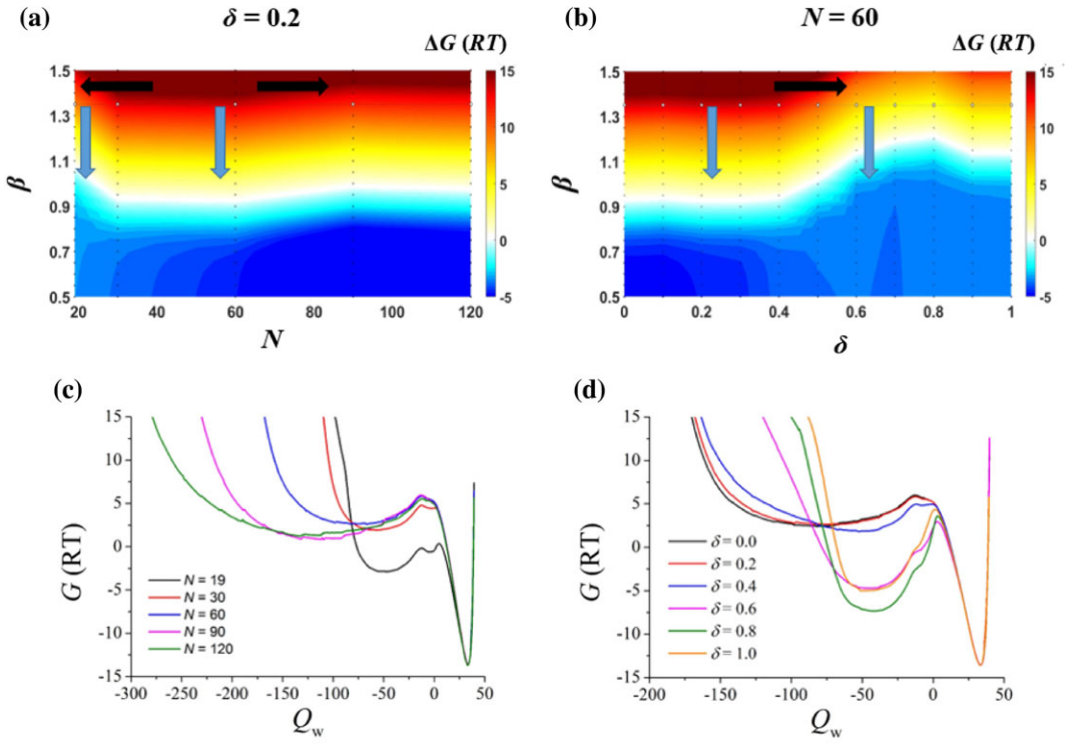

Figure 3. The effect of disordered linkers, taking 1BRS as an example. (a) The color map of the free energy change, $\Delta G_{\text{Closed–Open}}$, as a function of the disordered linker's length, N, and the domain–domain attraction, β, when the linker was in a coiled state (δ = 0.2). The white color represents $\Delta G_{\text{Closed–Open}}$ = 0 and the boundary of the closed-open transition. The black arrows show the effect due to the linker's length, N, while the blue arrows show the effect of decreasing β or allosteric binding affinity. β was equal to 1.35 for wild type Barnase/Barstar. (b) $\Delta G_{\text{Closed–Open}}$ as a function of the linker's flexibility parameter, δ, and the domain–domain attraction, β, while the linker length was fixed at N = 60. The coil-globule transition of the linker occurred in a rough range of $\delta \in$ [0.4, 0.6]. (c, d) The free energy profiles as a function of the binding reaction coordinate, Q_w (positive for the bound state and negative for the unbound state with $-Q_w$ approximately proportional to the domain separation distance), at β = 1.35.

$$\Delta G_{\text{Closed-open}} = -RT\ln\frac{[\text{A}\sim\text{B}]}{[\widehat{\text{AB}}]} \quad (9)$$

Here [.] represents the equilibrium concentration of the corresponding species. From Equation 9, it can be seen that the open-closed ratio of the with-linker system is independent of the total protein concentration $c_{\text{total}} = [\text{A}\sim\text{B}] + [\widehat{\text{AB}}]$. However, for the without-linker system, the open–closed (unbound/bound) ratio of the domain A or B, that is, [A]/[AB] or [B]/[AB], is influenced by the total protein concentrations $[\text{A}]_{\text{total}}$ = [A] + [AB] and $[\text{B}]_{\text{total}}$ = [B] + [AB] in the system. To measure the effect of disordered linker in terms of the characteristic concentration of the without-linker system where the open–closed ratios in two systems are equal, that is, the induced local concentration by the linker, we consider a simplified situation where B is excess over A. Under such a case, [B] ≈ $[\text{B}]_{\text{total}}$ and Equation 8 is simplified into a first-order reaction for A to give

$$\frac{[\text{A}]}{[\text{AB}]} = \frac{K_{\text{d}}}{[\text{B}]_{\text{total}}} \quad (10)$$

Now, the open–closed ratio of domain A is not influenced by its total concentration, $[\text{A}]_{\text{total}}$, being similar to the with-linker system. It is noted that the ratio is dependent on $[\text{B}]_{\text{total}}$. When the open–closed ratios of the without-linker and with-linker systems equal to each other, that is,

$$\frac{[\text{A}]}{[\text{AB}]} = \frac{[\text{A}\sim\text{B}]}{[\widehat{\text{AB}}]} \quad (11)$$

$[\text{B}]_{\text{total}}$ is solved to be

$$[\text{B}]_{\text{total}} = [\text{B}]_0 \equiv K_{\text{d}}\exp\left(\frac{\Delta G_{\text{closed-open}}}{RT}\right). \quad (12)$$

Here we use a different symbol $[\text{B}]_0$ to emphasize that it is the $[\text{B}]_{\text{total}}$ value at which the open-closed ratios of the two systems are same while $[\text{B}]_{\text{total}}$ is generally regarded as a free variable of the system.

K_{d} in the without-linker system is mainly determined by the protein–protein interaction. When a flexible linker is added to connect two domains, the same protein–protein (domain–domain) interaction contributes to $\Delta G_{\text{Closed–Open}}$ of the with-linker system together with the influence from the linker. Interestingly, the contribution of the domain–domain interaction to K_{d} cancels out its contribution to $\Delta G_{\text{Closed–Open}}$ in Equation 12, resulting in a characteristic concentration $[\text{B}]_0$ nearly independent on the interaction strength. For example, with our reweighting method, when tuning protein–protein interaction strength β, the curves of $\Delta G_{\text{Closed–Open}}$ and $-\ln K_{\text{d}}$ as a function of β experience the same tendency, see Figure 4.

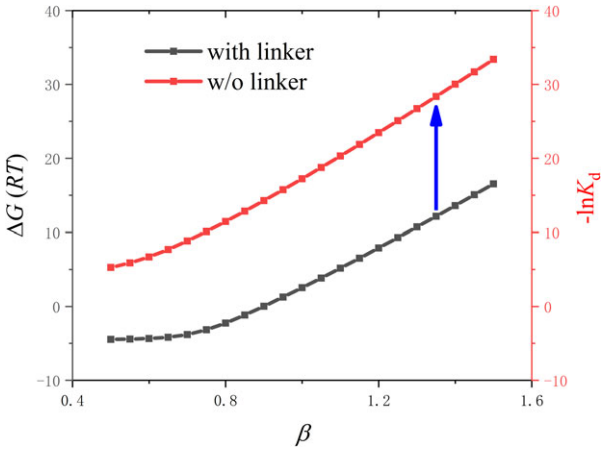

Figure 4. Reweighting curves of $\Delta G_{\text{Closed–Open}}$ and $-\ln K_{\text{d}}$ as a function of β, taking 1BRS for example. In the without-linker system, $-\ln K_{\text{d}}$ as a function of protein–protein interaction strength β is shown as red curve. In the with-linker system, $\Delta G_{\text{Closed–Open}}$ as a function of β is shown as black curve. For the wild type, Barnase/Barstar, β = 1.35. The blue arrow indicates the stable intercept difference between them.

In the applicable range of binding affinity, $\ln K_{\text{d}}$ is linear to β:[39]

$$-\ln K_{\text{d}} = k\beta + a_1. \quad (13)$$

For the with-linker system, a similar linear relationship is also found:

$$\frac{\Delta G_{\text{closed-open}}}{RT} = k\beta + a_2. \quad (14)$$

The $k\beta$ term in Equations 13 and 14 comes from the domain–domain interaction energy.[39] Then $[\text{B}]_0$ defined in Equation 12 can be expressed as:

$$[B]_0 = \exp(a_2 - a_1), \quad (15)$$

which is independent on β.

Therefore, $[\text{B}]_0$ can be used to quantitatively measure the effect of the linker in connecting two domains, which can be readily determined from simulated results. $[\text{B}]_0$ interplays with the domain–domain interaction (characterized by the binding affinity K_{d} of the without-linker system) to determine $\Delta G_{\text{Closed–Open}}$ of the with-linker system and the allosteric pathways: when $K_{\text{d}} < [\text{B}]_0$, $\Delta G_{\text{Closed–Open}} > 0$ and the closed state is preferred, so an auto-inhibition mechanism described in the left panels in Figure 1 is applicable for allosteric regulation; when $K_{\text{d}} > [\text{B}]_0$, $\Delta G_{\text{Closed–Open}} < 0$ and the open state is preferred to enable an affinity-enhancing mechanism for allosteric regulation in the right panels in Figure 1.

An alternative way to consider the local concentration

There is an alternative way to analyze the local concentration caused by adding the linker (Fig. 5). The

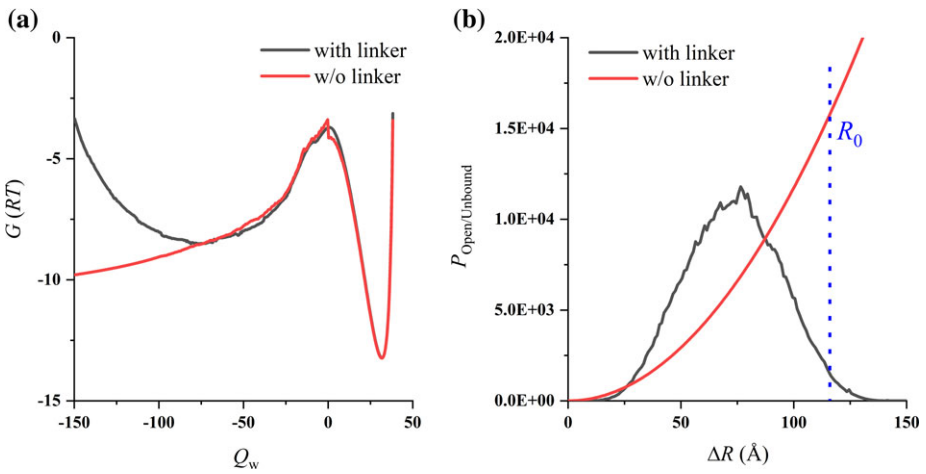

Figure 5. An alternative way to analyze the local concentration, taking 1BRS for example. (a) The free energy profiles: with (black) and without (red) the linker. (b) The population distribution as a function of the mass-center distance between two domains (ΔR): the open (black) and unbound (red) state. β = 1.0 was used in simulation.

introduction of a linker had little influence on the bound (closed) state, so the basins at $Q_w > 0$ in the free energy profile could be aligned to be well overlapped with that without a linker [Fig. 5 (a)]. However, it had a marked effect on the unbound (open) state with $Q_w < 0$ [Fig. 5(a)]: with the linker, the two domains were restricted to be near each other; without the linker, the two domains could separate at any distance, and the free energy decreases with the distance because of the increased space volume. The difference is demonstrated more clearly when we plotted the population distribution of the unbound (open) state as a function of the mass-center distance between two domains (ΔR) [Fig. 5 (b)]. Based on the $P(\Delta R)$ curves for the systems with and without a linker, a characteristic distances (R_0), is defined as the distance where the cumulative population $\int_0^{R_0} P(\Delta R)d\Delta R$ of the system with and without a linker is equal [Fig. 5(b)]. In other words, if one protein A molecule and one protein B molecule were put into a spherical volume with a radius of R_0 and A (or B) was fixed at the sphere center, the resulting unbound–bound ratio would be equal to the open–closed ratio when a linker is added. The determined vale of R_0 does not depend on the domain–domain attraction strength since the relative population between the unbound state and the open is note affected by the potential-well depth of the bound and the closed states. R_0 can be further converted into the corresponding characteristic concentration as

$$c_{R_0} = \frac{1}{N_A * \frac{4}{3}\pi R_0^3}, \quad (16)$$

where N_A is Avogadro's number.

Numerical results on local concentration

To measure the effect of disordered linker, the characteristic local concentration $[B]_0$ and c_{R0} were determined from simulations for all four protein systems listed in Table I, and the results were shown in Figure 6 and Table II when the linker length is N = 60. Although defined in different ways, $[B]_0$ and c_{R0} show no significant difference. The four protein systems are markedly different in structure order/disorder, binding affinity, surface/interface and the end-to-end distance R_{ee} in the complex, but the resulting characteristic local concentrations lie in a narrow range of 0.2–0.6 mM. It indicated that the introduction of a disordered linker would fix the effective concentration at an order of magnitude of mM. As a comparison, the values of $[B]_0$ evaluated from the experimental[29,54–57] K_d and $\Delta G_{Closed-Open}$ were among 0.2–5.3 mM, for different linker length among 8–44 residues. When the linker-domain excluded volume interaction is increased, $[B]_0$ would decreases slightly due to the enhanced entropic effect (Supporting Information Fig. S6).

This provides a clue in understanding the benefit of disordered linker in allosteric regulation. If a pair of protein domains have a low expression level, for example 10 nM, the binding probability is very small if K_d = 10 μM. Introducing a flexible linker will significantly enhance the local concentration and greatly promote the binding probability, making the regulation of weakly interacting possible.

The largest $[B]_0$ difference in Figure 6 occurs between 1BRS (0.265 mM) and 1A0O (0.509 mM), giving a ratio of 2.4, although their R_{ee} is almost identical (48 Å vs. 49 Å). It suggested that the end-to-end distance in native structure was not so important as the protein scaffold. To investigate the influence of other factors, $[B]_0$ of 1BRS was determined under

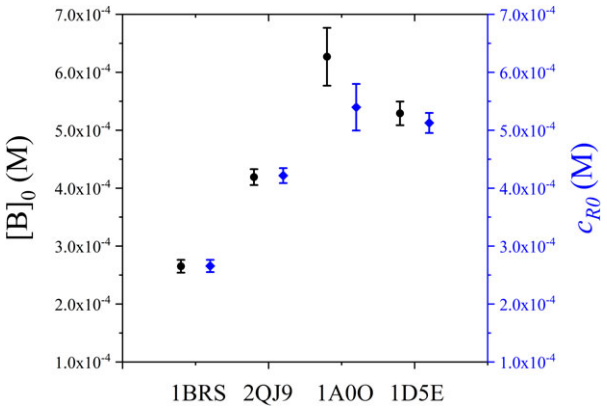

Figure 6. Characteristic concentration of disordered linker (N = 60 with δ = 0.2) in four protein systems. The effective local concentration $[B]_0$ and c_{R0} were shown as black circles and blue diamonds, respectively.

various linker length (N) and flexibility (δ), see Figure 7. Remarkably, a coil-to-globule transition of a N = 60 disordered linker brought 3 order of magnitude of decreasing to $[B]_0$, from mM to μM, see Figure 7(a). Highly collapsed conformations of linker severely prevented the domains from binding, resulting in the weakening of local concentration increasing effect. In realistic conditions, such a coil-to-globule transition is difficult to achieve.[46] On the other hand, the linker length is a more realistic parameter to control, which caused 2 order of magnitude to the variation of $[B]_0$ in our simulations, see Figure 7(b). When the linker is too short, elastic stress is induced in the closed state to reduce its stability and to decrease the effective $[B]_0$. When the linker is too long, the open state has much more allowed conformations, leading to the decreasing of $[B]_0$. For example, when N increases from 60 to 120, $[B]_0$ decreases with almost one order of magnitude. Therefore, $[B]_0$ is mainly determined by the linker while little affected by the domains.

$[B]_0$ is a bridge in relating K_d of the without-linker protein system to $\Delta G_{Closed–Open}$ of the with-linker system, which would be useful in experimental study. Experimentally, usually only K_d or $\Delta G_{Closed–Open}$ was measured, while it is very rare to obtain both K_d and $\Delta G_{Closed–Open}$ for the same domain pairs probably due to experimental obstacle. For example, the most favorable $\Delta G_{Closed–Open}$ range to be directly measured is around $\Delta G_{Closed–Open}$ = 0, but this would correspond to K_d at an order of magnitude near mM, which is difficult to be measured in many usual techniques. With the approximate value $[B]_0$ from simulations, it is thus straightforward to measure $\Delta G_{Closed–Open}$ and convert it into K_d based on Equation 15, or vice versa. It would be helpful in evaluate the potential of engineering an existing protein system by removing or adding the connecting disorder linker.

Comment on possible improvement on the model

The open/close equilibrium of multidomain proteins is determined by the competition between the entropic stabilization of open state (destabilization of closed state) and the enhancement of local effective concentration to favor the closed state. In our study, we adopted a Gō-like coarse-grained model considering only C_α coordinates, which is both concise and computationally efficient. This may underestimate the entropic contribution and is worthy of further study, for example, using atomic models. Furthermore, to establish a conceptual model, we did not consider specific intralinker or linker–domain interaction which involved sequence specificity. In realistic situations, the physical properties of disordered linker were not limited to the chain length and flexibility. Many other specific properties are also important. For example, positively and negatively charge residues of disordered linkers were shown to be important to the conformation distributions.[58,59] Asphericity was also used as another important factor of IDRs.[60] Moreover, phosphorylation sites, polar or hydrophobic residues, even directed-proline residues could affect the dimension of IDRs.[61] The attracting interaction between domains and linker may stabilize the closed state and increase $[B]_0$, while the enhanced excluded-volume (repulsive) interaction between domains and linker would decrease $[B]_0$. The incorporation of these effects is straightforward by adopting other more delicate coarse-grained models, for example, the modified Gō-like model with sequence-dependent hydrophobic interaction,[45,62] the SIRAH model with incorporated ion and coulomb interaction,[63,64] or Miyazawa–Jernigan-type contact potentials.[65,66]

Table II. Summary of Free Energy and Effective Concentration Results[a]

PDB	β	K_d (M)	ΔG (RT)	$[B]_0$ (M)	c_{R0} (M)
1BRS	1.0	2.10E-5 ± 8.69E-7	2.54	2.65E-4 ± 1.10E-5	2.66E-4 ± 1.06E-5
2QJ9	1.0	4.11E-8 ± 1.34E-9	9.23	4.19E-4 ± 1.37E-5	4.22E-4 ± 1.29E-5
1A0O	1.0	6.71E-3 ± 5.34E-4	−2.37	5.09E-3 ± 2.79E-4	5.40E-4 ± 4.02E-5
1D5E	1.0	8.10E-4 ± 3.13E-5	−0.43	5.29E-4 ± 2.04E-5	5.13E-4 ± 1.73E-5

[a]The data were derived from simulation of disordered linker (N = 60 with δ = 0.2) in four protein systems. The domain–domain binding affinity parameter β was set to 1.0 to align different protein systems. K_d and ΔG were calculated by free energy profiles as Figure 2. $[B]_0$ and c_{R0} was calculated as Figures 4 and 5, and plotted in Figure 6.

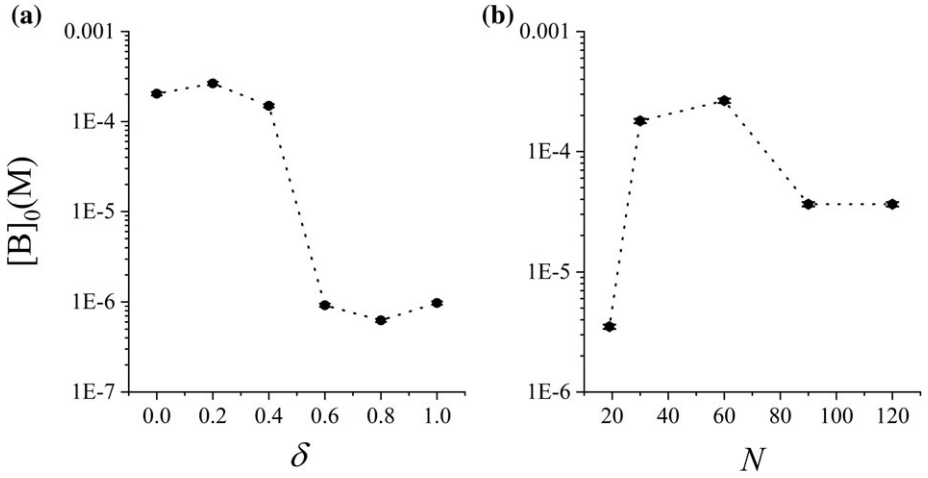

Figure 7. Characteristic concentration $[B]_0$ of disordered linker in 1BRS as a function of (a) the linker flexibility δ and (b) the linker length N. N = 60 in (a) and δ = 0.2 in (B).

Conclusion

In summary, we investigated the role of a disordered linker in stabilizing the open or the closed states of multidomain allosteric proteins. Coarse-grained molecular dynamics simulation was performed to determine the free energy profile of the system under various conditions. Deduction based on a thermodynamic analysis was conducted to reveal that the influence of linker can be fully characterized by an effective local concentration $[B]_0$. When $K_d < [B]_0$, the closed state would be favored to enable an auto-inhibition mechanism for allosteric regulation; while the open state would be preferred to enable an affinity-enhancing mechanism for allosteric regulation when $K_d > [B]_0$. We used four protein systems with markedly different domain–domain binding affinity and structural order/disorder as model systems to investigate the relationship between $[B]_0$ and its various determinants. $[B]_0$ is nearly independent of K_d, while the linker length N is its main practical determinant. $[B]_0$ of a linker with N = 40–60 was determined to be in a narrow range of 0.2–0.6 mM, while a too short or too long length would dramatically decrease $[B]_0$. Overall, the introduction of a flexible linker significantly increases the possibility of the closed state, making the regulation of weakly interacting partners possible.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant 21633001), and the Ministry of Science and Technology of China (grant 2015CB910300). We thank Prof. Peter Tompa, Prof. Richard Kriwacki, Prof. Shinichi Tate and Prof. Zhuqing Zhang for helpful discussions.

References

1. Monod J, Wyman J, Changeux JP (1965) On nature of allosteric transitions - a plausible model. J Mol Biol 12: 88–118.
2. Koshland DE, Nemethy G, Filmer D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry 5:365–385.
3. Petit CM, Zhang J, Sapienza PJ, Fuentes EJ, Lee AL (2009) Hidden dynamic allostery in a pdz domain. Proc Natl Acad Sci U S A 106:18249–18254.
4. Popovych N, Tzeng S-R, Tonelli M, Ebright RH, Kalodimos CG (2009) Structural basis for camp-mediated allosteric control of the catabolite activator protein. Proc Natl Acad Sci USA 106:6927–6932.
5. Ma XM, Meng H, Lai LH (2016) Motions of allosteric and orthosteric ligand-binding sites in proteins are highly correlated. J Chem Inform Model 56:1725–1733.
6. Popovych N, Sun SJ, Ebright RH, Kalodimos CG (2006) Dynamically driven protein allostery. Nat Struct Mol Biol 13:831–838.
7. Tzeng SR, Kalodimos CG (2009) Dynamic activation of an allosteric regulatory protein. Nature 462:368–372.
8. Gunasekaran K, Ma BY, Nussinov R (2004) Is allostery an intrinsic property of all dynamic proteins? Proteins 57:433–443.
9. Nussinov R (2016) Introduction to protein ensembles and allostery. Chem Rev 116:6263–6266.
10. Tsai C-J, Nussinov R (2014) A unified view of "how allostery works". Plos Comput Biol 10:e1003394.
11. Liu ZR, Huang YQ (2014) Advantages of proteins being disordered. Protein Sci 23:539–550.
12. Tantos A, Han KH, Tompa P (2012) Intrinsic disorder in cell signaling and gene transcription. Mol Cell Endocrin 348:457–465.
13. Tompa P (2014) Multisteric regulation by structural disorder in modular signaling proteins: An extension of the concept of allostery. Chem Rev 114:6715–6732.
14. Guo J, Zhou H-X (2016) Protein allostery and conformational dynamics. Chem Rev 116:6503–6515.
15. Papaleo E, Saladino G, Lambrughi M, Lindorff-Larsen K, Gervasio FL, Nussinov R (2016) The role of protein loops and linkers in conformational dynamics and allostery. Chem Rev 116:6391–6423.
16. Wallin S (2017) Intrinsically disordered proteins: Structural and functional dynamics. Res Rep Biol 8:7–16.
17. Motlagh HN, Wrabl JO, Li J, Hilser VJ (2014) The ensemble nature of allostery. Nature 508:331–339.
18. Yun TKX, Qin T, Liu Y, Lai LH (2016) Discovery of non-atp-competitive inhibitors of polo-like kinase1. Chemmedchem 11:713–717.

19. Yun T, Qin T, Liu Y, Lai L (2016) Identification of acylthiourea derivatives as potent plk1 pbd inhibitors. Eur J Med Chem 124:229–236.
20. Qin T, Chen F, Zhuo X, Guo X, Yun T, Liu Y, Zhang C, Lai L (2016) Discovery of novel polo-like kinase 1 polo-box domain inhibitors to induce mitotic arrest in tumor cells. J Med Chem 59:7089–7096.
21. Wang J, Tochio N, Kawasaki R, Tamari Y, Xu N, Uewaki J, Utsunomiya-Tate N, Tate S (2015) Allosteric breakage of the hydrogen bond within the dual-histidine motif in the active site of human pin1 pplase. Biochemistry 54:5242–5253.
22. Guo J, Pang X, Zhou H-X (2015) Two pathways mediate interdomain allosteric regulation in pin1. Structure 23: 237–247.
23. Kissinger CR, Parge HE, Knighton DR, Lewis CT, Pelletier LA, Tempczyk A, Kalish VJ, Tucker KD, Showalter RE, Moomaw EW, Gastinel LN, Habuka N, Chen XH, Maldonado F, Barker JE, Bacquet R, Villafranca JE (1995) Crystal-structures of human calcineurin and the human fkbp12-fk506-calcineurin complex. Nature 378:641–644.
24. Yang J, Ten Eyck LF, Xuong NH, Taylor SS (2004) Crystal structure of a camp-dependent protein kinase mutant at 1.26 angstrom: New insights into the catalytic mechanism. J Mol Biol 336:473–487.
25. Cotelesage JJH, Puttick J, Goldie H, Rajabi B, Novakovski B, Delbaere LTJ (2007) How does an enzyme recognize co2? Int J Biochem Cell Biol 39: 1204–1210.
26. Xu WQ, Doshi A, Lei M, Eck MJ, Harrison SC (1999) Crystal structures of c-src reveal features of its autoinhibitory mechanism. Mol Cell 3:629–638.
27. Brown NR, Noble MEM, Lawrie AM, Morris MC, Tunnah P, Divita G, Johnson LN, Endicott JA (1999) Effects of phosphorylation of threonine 160 on cyclin-dependent kinase 2 structure and activity. J Biol Chem 274:8746–8756.
28. Zhou HX (2001) Single-chain versus dimeric protein folding: Thermodynamic and kinetic consequences of covalent linkage. J Am Chem Soc 123:6730–6731.
29. Zhou HX (2001) The affinity-enhancing roles of flexible linkers in two-domain DNA-binding proteins. Biochemistry 40:15069–15073.
30. Zhou HX (2001) Loops in proteins can be modeled as worm-like chains. J Phys Chem B 105:6763–6766.
31. Zhang J, Li WF, Wang J, Qin M, Wu L, Yan ZQ, Xu WX, Zuo GH, Wang W (2009) Protein folding simulations: From coarse-grained model to all-atom model. LUBMB Life 61:627–643.
32. Pincus DL, Cho SS, Hyeon CB, Thirumalai D (2008) Minimal models for proteins and RNA: From folding to function. Prog Mol Biol Translat Sci 84: 203–250.
33. Chan HS, Zhang ZQ, Wallin S, Liu ZR (2011) Cooperativity, local-nonlocal coupling, and nonnative interactions: Principles of protein folding from coarse-grained models. Annu Rev Phys Chem 62:301–326.
34. Hills RD, Brooks CL (2009) Insights from coarse-grained go models for protein folding and dynamics. Int J Mol Sci 10:889–905.
35. Xie W, Xu Z-R, Wang M, Xu S-C (2016) Molecular dynamics simulation for levo-benzedrine to transmit through molecular channels within D3R. Acta Phys Chim Sinica 32:907–920.
36. Huang YQ, Liu ZR (2009) Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: A critical assessment of the "fly-casting" mechanism. J Mol Biol 393:1143–1159.
37. Li MD, Sun TL, Jin F, Yu DQ, Liu ZR (2016) Dimension conversion and scaling of disordered protein chains. Mol Biosyst 12:2932–2940.
38. Kim YC, Hummer G (2008) Coarse-grained models for simulations of multiprotein complexes: Application to ubiquitin binding. J Mol Biol 375:1416–1433.
39. Cao HQ, Huang YQ, Liu ZR (2016) Interplay between binding affinity and kinetics in protein-protein interactions. Proteins 84:920–933.
40. Huang YQ, Liu ZR (2010) Smoothing molecular interactions: The "kinetic buffer" effect of intrinsically disordered proteins. Proteins 78:3251–3259.
41. Huang YQ, Liu ZR (2013) Do intrinsically disordered proteins possess high specificity in protein-protein interactions? Chemistry 19:4462–4467.
42. Qi Y, Wang Q, Tang B, Lai L (2012) Identifying allosteric binding sites in proteins with a two-state g(o) over-bar model for novel allosteric effector discovery. J Chem Theory Comput 8:2962–2971.
43. Badasyan A, Liu ZR, Chan HS (2008) Probing possible downhill folding: Native contact topology likely places a significant constraint on the folding cooperativity of proteins with similar to 40 residues. J Mol Biol 384:512–530.
44. Kaya H, Chan HS (2003) Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: How adequate is native-centric topological modeling? J Mol Biol 326:911–931.
45. Zarrine-Afsar A, Wallin S, Neculai AM, Neudecker P, Howell PL, Davidson AR, Chan HS (2008) Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. Proc Natl Acad Sci USA 105:9999–10004.
46. Li MD, Liu ZR (2016) Dimensions, energetics, and denaturant effects of the protein unstructured state. Protein Sci 25:734–747.
47. Yoo TY, Meisburger SP, Hinshaw J, Pollack L, Haran G, Sosnick TR, Plaxco K (2012) Small-angle x-ray scattering and single-molecule fret spectroscopy produce highly divergent views of the low-denaturant unfolded state. J Mol Biol 418:226–236.
48. Ziv G, Haran G (2009) Protein folding, protein collapse, and tanford's transfer model: Lessons from single-molecule fret. J Am Chem Soc 131:2942–2947.
49. Watkins HM, Simon AJ, Sosnick TR, Lipman EA, Hjelm RP, Plaxco KW (2015) Random coil negative control reproduces the discrepancy between scattering and fret measurements of denatured protein dimensions. Proc Natl Acad Sci USA 112:6631–6636.
50. Schreiber G, Fersht AR (1993) Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. Biochemistry 32:5145–5150.
51. Umulis D, O'Connor MB, Blair SS (2009) The extracellular regulation of bone morphogenetic protein signaling. Development 136:3715–3728.
52. Stewart RC, Van Bruggen R (2004) Association and dissociation kinetics for chey interacting with the p2 domain of chea. J Mol Biol 336:287–301.
53. Goldberg JM, Baldwin RL (1998) Kinetic mechanism of a partial folding reaction. 1. Properties of the reaction and effects of denaturants. Biochemistry 37:2546–2555.
54. Jana R, Hazbun TR, Fields JD, Mossing MC (1998) Single-chain lambda Cro repressors confirm high intrinsic dimer-DNA affinity. Biochemistry 37:6446–6455.
55. Klemm JD, Pabo CO (1996) Oct-1 pou domain DNA interactions: Cooperative binding of isolated subdomains and effects of covalent linkage. Genes Devel 10: 27–36.
56. Simoncsits A, Chen JQ, Percipalle P, Wang SL, Toro I, Pongor S (1997) Single-chain repressors containing

engineered DNA-binding domains of the phage 434 repressor recognize symmetric or asymmetric DNA operators. J Mol Biol 267:118–131.
57. Bain DL, Ackers GK (1994) Self-association and DNA-binding of lambda ci repressor n-terminal domains reveal linkage between sequence-specific binding and the c-terminal cooperativity domain. Biochemistry 33: 14679–14689.
58. Das RK, Pappu RV (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc Natl Acad Sci USA 110:13392–13397.
59. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. Proc Natl Acad Sci U S A 107:8183–8188.
60. Song J, Gomes G-N, Gradinaru CC, Chan HS (2015) An adequate account of excluded volume is necessary to infer compactness and asphericity of disordered proteins by forster resonance energy transfer. J Phys Chem B 119:15191–15202.
61. Martin EW, Holehouse AS, Grace CR, Hughes A, Pappu RV, Mittag T (2016) Sequence determinants of the conformational properties of an intrinsically disordered protein prior to and upon multisite phosphorylation. J Am Chem Soc 138:15323–15335.
62. Zhang Z, Chan HS (2010) Competition between native topology and nonnative interactions in simple and complex folding kinetics of natural and designed proteins. Proc Natl Acad Sci U S A 107:2920–2925.
63. Darre L, Rodrigo Machado M, Febe Brandner A, Carlos Gonzalez H, Ferreira S, Pantano S (2015) SIRAH: A structurally unbiased coarse-grained force field for proteins with aqueous solvation and long-range electrostatics. J Chem Theory Comput 11:723–739.
64. Machado MR, Pantano S (2016) SIRAH tools: Mapping, backmapping and visualization of coarse-grained models. Bioinformatics 32:1568–1570.
65. Miyazawa S, Jernigan RL (1996) Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. J Mol Biol 256:623–644.
66. Thomas PD, Dill KA (1996) An iterative method for extracting energy-like quantities from protein structures. Proc Natl Acad Sci USA 93: 11628–11633.