# Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins

**Vincent J. Hilser* and E. Brad Thompson**

Department of Biochemistry and Molecular Biology, and Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, TX 77555-1068

Edited by Robert L. Baldwin, Stanford University Medical Center, Stanford, CA, and approved February 21, 2007 (received for review January 12, 2007)

**Transcription factors and other allosteric cell signaling proteins contain a disproportionate number of domains or segments that are intrinsically disordered (ID) under native conditions. In many cases folding of these segments is coupled to binding with one or more of their interaction partners, suggesting that intrinsic disorder plays an important functional role. Despite numerous hypotheses for the role of ID domains in regulation, a mechanistic model has yet to be established that can quantitatively assess the importance of intrinsic disorder for intramolecular site-to-site communication, the hallmark property of allosteric proteins. Here, we present such a model and show that site-to-site allosteric coupling is maximized when intrinsic disorder is present in the domains or segments containing one or both of the coupled binding sites. This result not only explains the prevalence of ID domains in regulatory proteins, it also calls into question the classical mechanical view of energy propagation in proteins, which predicts that site-to-site coupling would be maximized when a well defined pathway of folded structure connects the two sites. Furthermore, in showing that the coupling mechanism conferred by intrinsic disorder is robust and independent of the network of interactions that physically link the coupled sites, unique insights are gained into the energetic ground rules that govern site-to-site communication in all proteins.**

allostery | ensemble | regulation | site-to-site communication

Over the past decade, the paradigm that proteins function by adopting highly ordered structures has been challenged by the observation that thousands of different proteins are likely to be intrinsically disordered (i.e., sample multiple conformations) or have intrinsically disordered (ID) domains under native conditions (1–8). Of particular significance is the growing body of evidence that intrinsic disorder is found in disproportionately higher amounts in cell signaling proteins and transcription factors, suggesting an important role in their regulatory capacity (5). Indeed for cases where detailed study has been performed, structure formation in these proteins (or domains) is linked to ligand binding in other parts of the molecule, indicating that the order/disorder transition is coupled to long-range allosteric communication within the molecule and is therefore important to its functional role (5). Interestingly, within specific classes of regulatory proteins [such as the steroid hormone receptors (9)], functionally analogous domains relevant to transcription regulation appear to be ID in each member, even though there is little sequence conservation in these regions. This finding suggests that the underlying regulatory mechanism is both effective and robustly encoded in nature.

Hypotheses for the role of intrinsic disorder include: high-specificity/low-affinity binding (3, 5), rapid protein turnover (1), and high specificity for multiple targets (1–5). However, despite the clear experimental evidence for the existence of ID domains within regulatory proteins (3, 5, 7), a mechanistic model that provides a quantitative rationale for their presence has yet to be established. Because the common thread that connects most regulatory proteins is that their function is modulated by allosteric coupling to other sites, we set out to investigate whether having ID domains could offer a protein a selective advantage in coupling two sites. Here, a general theory is presented for the role of ID domains in mediating allosteric interactions, and it is demonstrated, via an unbiased sampling of parameter space within the context of this model, that intrinsic disorder optimizes allosteric coupling.

## Results

**Theory.** It is well established that proteins have modular structure (10), and that multidomain regulatory proteins often segregate the binding sites for each ligand into the different structural domains (9). Accordingly, an allosteric protein can be represented, to a first approximation, as a group of interacting domains, the simplest of which is the two-domain protein shown in Fig. 1*A*. We will use this description to investigate allostery. Because we want to explore the role of intrinsic disorder on allosteric control, each domain will be allowed the freedom to be independently folded or unfolded, resulting in four possible states (i.e., N, 1, 2, and U), representing all combinations of having each domain unfolded. The important feature of this representation, which is the reason that each domain can "sense" the other, is that the energy of each state (relative to the N state) is composed of the free energy of unfolding each domain plus the energy of breaking the interactions between them ($\Delta g_{\text{int}}$). For the simple system shown in Fig. 1, the partition function, $Q$, which is the sum of the statistical weights of all of the states, is

$$Q = 1 + K_{\text{II}}\phi_{\text{int}} + K_{\text{I}}\phi_{\text{int}} + K_{\text{I}}K_{\text{II}}\phi_{\text{int}}, \qquad [1]$$

where $K_{\text{II}} = \exp(-\Delta G_{\text{II}}/RT)$, $K_{\text{I}} = \exp(-\Delta G_{\text{I}}/RT)$, and $\phi_{\text{int}} = \exp(-\Delta g_{\text{int}}/RT)$. The corresponding probabilities are just the statistical weights normalized to the partition function, as shown in the last column of Fig. 1*A*.

Coupling between domains results when $\Delta g_{\text{int}} \neq 0$ (i.e., $\phi_{\text{int}} \neq 1$). Although $\Delta g_{\text{int}}$ can arise from any source, associating this parameter with a physically meaningful process adds clarity to the model. For cases where $\Delta g_{\text{int}}$ is positive, it is energetically unfavorable to break the interaction. Such a situation would exist, for example, with two complementary hydrophobic surfaces, wherein it would be energetically more favorable to interact with each other than with solvent. For situations where $\Delta g_{\text{int}}$ is negative, on the other hand, exactly the opposite effect would be observed; interaction of the surfaces with solvent would be more favorable than the interaction with each other. Such a situation might exist for the same hydrophobic surfaces described above, but at low temperatures (11). We note however, and will discuss below, that the physical basis for the interaction is not relevant, only the magnitude of the value.

Author contributions: V.J.H. and E.B.T. designed research; V.J.H. performed research; V.J.H. analyzed data; and V.J.H. and E.B.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: ID, intrinsically disordered; CR, coupling response.

*To whom correspondence should be addressed. E-mail: vjhilser@utmb.edu.

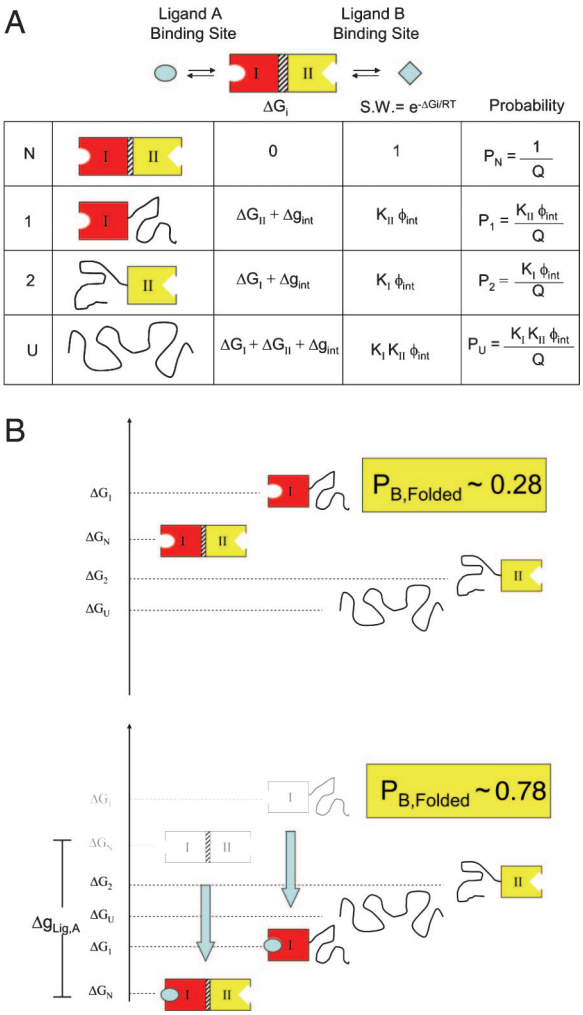


**Fig. 1.** Model for allosteric coupling. (*A*) Schematic representation of a test model for allosteric coupling. Shown is a hypothetical two-domain protein that can bind two different ligands (A and B), one in each domain. Each domain can be folded or unfolded, resulting in four possible states (i.e., N, 1, 2, and U). (*B*) Site-to-site coupling is evaluated through the addition of ligand A, which binds to only domain I. Because ligand A stabilizes those states that bind A (blue arrows), the ensemble probabilities are redistributed. Depending on the relative probabilities of each state before the addition of ligand A, those states that can bind ligand B will either be: (*i*) stabilized, (*ii*) destabilized, or (*iii*) not affected by the binding of ligand A. The effect can be quantified through a parameter called the CR [$CR = \Delta(P_N + P_2)/\Delta \ln Z_{Lig,A}$], which is the change in the probability of states that can bind ligand B (i.e., $P_{B,Folded} = P_N + P_2$) as a result of binding ligand A. For this model, values for the intrinsic stabilities of each domain (i.e., $\Delta G_I$ and $\Delta G_{II}$) are relative to the folded conformation; positive values mean that the folded conformation of each domain is more stable, and negative values mean that the unfolded form is more stable. The interaction energy (i.e., $\Delta g_{int}$) is referenced to the interacting (i.e., N) state; positive values mean that it is more energetically favorable for the interaction to be formed than to have the domains interact with solvent, and negative values mean that it is more favorable to have the domains interact with solvent. The parameters used, $\Delta G_I = -2.3$ kcal/mol, $\Delta G_{II} = -0.7$ kcal/mol, $\Delta g_{int} = 1.6$ cal/mol, and $\Delta g_{Lig,A} = -3.0$ kcal/mol, result in a significant CR and is part of the positively coupled node of parameters in Fig. 2*A*. (*Upper*) Without ligand A. (*Lower*) With ligand A.

To explore the extent of coupling, a perturbation can be applied to either domain. In principle, a perturbation can be in the form of a mutation, the ionization or chemical modification of a residue, or the binding of a ligand. For the current case, where we want to explore allosteric coupling, the perturbation arises from the binding of a ligand. To facilitate this scenario, a binding site for ligand A is introduced into domain I, and a binding site for ligand B is introduced into domain II. We are interested in elucidating how the binding of ligand A to domain I can influence the ability of the protein to bind ligand B in domain II.

The binding of ligand by ID proteins is usually associated with folding of the disordered domain (1), indicating that the affinity for ligand of the folded conformation is greater than the affinity for ligand of the disordered conformation. To capture this observation, the model is constructed such that each domain can bind its putative ligand only when it is folded. Thus, state 1 can only bind ligand A, state 2 can only bind ligand B, state N can bind both ligands, and state U can bind neither ligand.

Because states N and 1 are able to bind ligand A, the partition function (Eq. **1**) in the presence of ligand A becomes

$$Q = Z_{\mathrm{Lig,A}}(1 + K_{\mathrm{II}}\phi_{\mathrm{int}}) + K_{\mathrm{I}}\phi_{\mathrm{int}} + K_{\mathrm{I}}K_{\mathrm{II}}\phi_{\mathrm{int}}, \qquad [2]$$

where $Z_{\mathrm{Lig,A}} = 1 + K_{\mathrm{a,A}}[A]$, and $K_{\mathrm{a,A}}$ is the intrinsic association constant of domain I for ligand A (we note that at concentrations of $[A] \ll 1/K_{\mathrm{a,A}}$, Eq. **2** reduces to Eq. **1**). As Eq. **2** reveals, adding ligand A to the system results in a redistribution of the ensemble probabilities. The question with regard to allostery is, what effect did the binding of ligand A (to states N and 1) have on the probability of the states that can bind ligand B (states N and 2)? The question can be resolved analytically by examining the expression for the combined probability of states N and 2 ($P_{\mathrm{B,Folded}}$) both without and with ligand A:

without ligand

$$P_{\mathrm{B,Folded}} = \frac{1 + K_{\mathrm{I}}\phi_{\mathrm{int}}}{1 + K_{\mathrm{II}}\phi_{\mathrm{int}} + K_{\mathrm{I}}\phi_{\mathrm{int}} + K_{\mathrm{I}}K_{\mathrm{II}}\phi_{\mathrm{int}}}, \qquad [3a]$$

with ligand

$$P_{\mathrm{B,Folded}} = \frac{Z_{\mathrm{Lig,A}} + K_{\mathrm{I}}\phi_{\mathrm{int}}}{Z_{\mathrm{Lig,A}}(1 + K_{\mathrm{II}}\phi_{\mathrm{int}}) + K_{\mathrm{I}}\phi_{\mathrm{int}} + K_{\mathrm{I}}K_{\mathrm{II}}\phi_{\mathrm{int}}}. \qquad [3b]$$

As the expressions indicate, the effect of $Z_{\mathrm{Lig,A}}$ on the probability of states in which the binding site for B is folded (and therefore competent to bind ligand B) will depend on the magnitudes of the statistical weights of the individual states, which in turn will depend on the intrinsic stabilities of each domain, $\Delta G_{\mathrm{I}}$ and $\Delta G_{\mathrm{II}}$, and the interaction energy between them, $\Delta g_{\mathrm{int}}$. This point is demonstrated in Fig. 1*B* with an arbitrary set of parameter values for $\Delta G_{\mathrm{I}}$, $\Delta G_{\mathrm{II}}$, and $\Delta g_{\mathrm{int}}$. For this example, the free energy of each state without ligand is determined as shown in Fig. 1*A*, and the corresponding probability for the states that bind ligand B (Eq. **3a**) are modest (i.e., $P_{\mathrm{B,Folded}} \approx 0.28$). Thus, domain I is almost never folded, and domain II is folded only 30% of the time. The addition of ligand A, however, stabilizes each state that binds ligand A by the amount

$$\Delta g_{\mathrm{Lig,A}} = -\mathrm{RT}\cdot\ln Z_{\mathrm{Lig,A}} = -RT\cdot\ln(1 + K_{\mathrm{a,A}}[A]), \qquad [4]$$

and the probabilities of each state are redistributed (Fig. 1*B*). For the particular example shown, the stabilization of states that bind ligand A results in a substantial shift in the probability of states that bind ligand B (Eq. **3b**). The physical basis of this coupling between sites is discussed in detail below. To quantify the coupling we define the allosteric coupling response (CR) as the degree to which the probability of states that can bind ligand B is affected for a given perturbation to states that can bind ligand A:

$$CR = \frac{\Delta P_{\mathrm{B,Folded}}}{\Delta \ln Z_{\mathrm{Lig,A}}} = \frac{|P_{\mathrm{B,Folded}}(A) - P_{\mathrm{B,Folded}}(A = 0)|}{\ln Z_{\mathrm{Lig,A}}} \qquad [5]$$

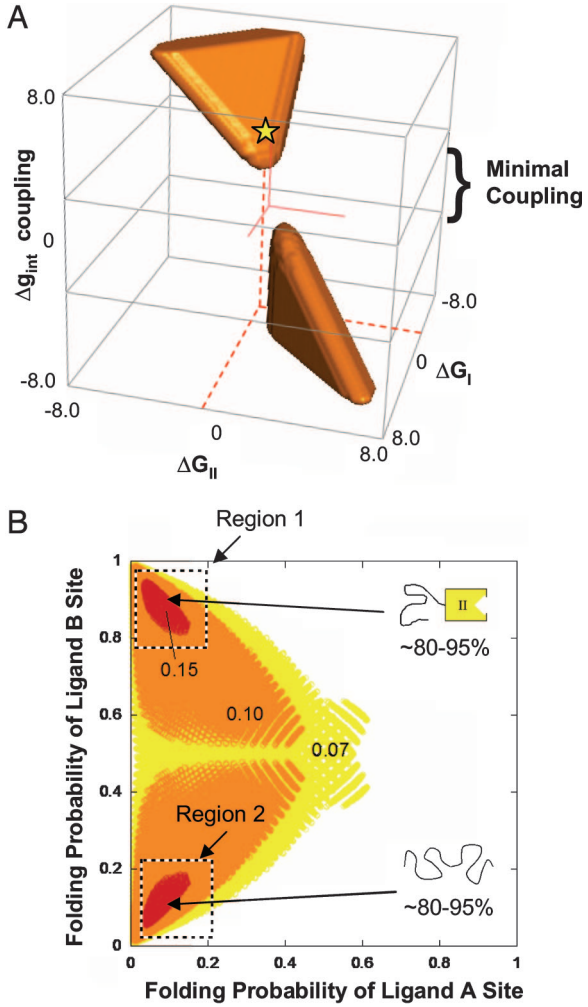


**Fig. 2.** Proteins with ID domains are optimized for allosteric coupling. (*A*) 3D plot showing all parameter combinations that generate a CR ≥ 0.10. A wide range of parameter combinations is sufficient to elicit a high response. The absence of points in the region marked minimal coupling indicates that although a wide range of parameter values can combine to produce a high CR, significant interaction energy (i.e., for this case, $|\Delta g_{int}| > \approx 1.0$ kcal/mol) is a prerequisite to coupling. All energies are in kcal/mol. The star and dashed red lines indicate those parameter values used to generate Fig. 1*B*, and solid red axes denote the origin. (*B*) Plot of the folding probability of the ligand A site (i.e., $P_N + P_1$) vs. the folding probability of the ligand B site (i.e., $P_N + P_2$) showing only those parameter combinations where the CR exceeds a specified response threshold. Three different response thresholds are shown: 0.07 (yellow), 0.10 (orange), and 0.15 (red). The maximum responses (dashed boxes) are observed in two regions. In region 1, domain I is unfolded and domain II is folded. Binding of ligand A folds domain I, but because of unfavorable domain coupling, domain II unfolds. In region II, both domains are unfolded. Binding of ligand A folds domain I, and as a result of favorable domain coupling, domain II folds.

In words, the CR (Eq. **5**) is a measure of the sensitivity of site B to perturbations (such as ligand binding) at site A, and thus provides a quantifiable metric of allosteric coupling potential. To determine which parameters maximize this potential, an unbiased search of parameter space can be performed by systematically exploring all possible combinations of values for $\Delta G_I$, $\Delta G_{II}$, and $\Delta g_{int}$ (Fig. 2).

**Results of the Model.** The classic view of allosteric coupling is that two sites are coupled through a network of interactions that extend throughout the protein and connect the two sites, in

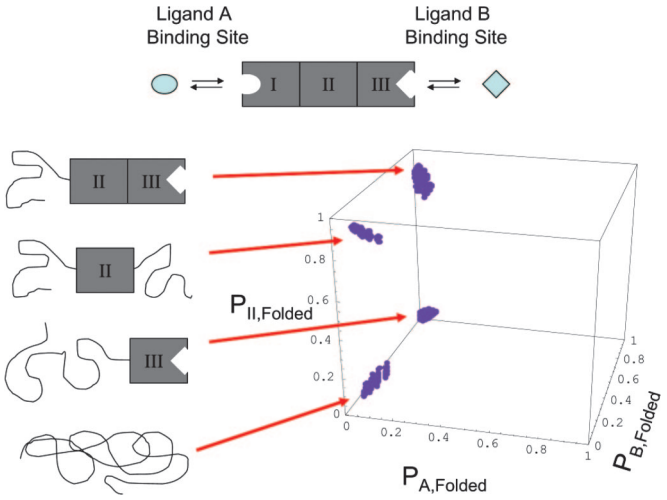


**Fig. 3.** Extension of the allosteric model to three domains. 3D plot of probability space for the three-domain protein (analogous to the 2D plot shown in Fig. 2*B* for the two-domain protein). Shown are those parameters that result in a CR ≥0.07 (purple points). Arrows denote which is the dominant state for each one of the four nodes (groups of points) observed.

essence, that there is an energetic pathway linking the sites (12). If this is the case, it might be expected that site-to-site coupling would be maximized when a well defined pathway of stable, folded structure connects the two sites. Paradoxically, such a conclusion is not borne out in the current analysis. Instead, an inverse relationship between allosteric coupling potential and the stability within the molecule is observed, and this relationship provides insight into the ground rules governing site-to-site coupling.

Initial inspection of the parameter space that maximizes coupling (Fig. 2*A*) reveals that there are distinct regions of energetic space within which different combinations of energetic values can facilitate coupling. Also evident is that there are two nodes for the two-domain protein. The origin of this behavior becomes apparent when the stabilities from Fig. 2*A* are converted to the probabilities of states that can bind ligand A and B. Shown in Fig. 2*B* are the parameter combinations that produce CR values in excess of 0.07 (yellow), 0.10 (orange), and 0.15 (red). Several observations can be made. First and most important, allosteric coupling is found to be maximized when the domains containing one or both binding sites are ID a significant fraction of the time in the absence of ligand, a result that is consistent with the prevalence of ID segments in, for example, transcription factors (5).

Second, although the individual energetic parameters that can facilitate coupling vary considerably (i.e., a unique set of parameter values for $\Delta G_I$, $\Delta G_{II}$, and $\Delta g_{int}$ are not necessary), in all cases significant interaction energy is required. This point is discussed in more detail below. We note that because all possible conditions were tested the result obtained here is not predetermined by the specifics of the analysis. Indeed, varying the degree of complexity of the model (e.g., introducing additional domains) does not affect the results. This is demonstrated in Fig. 3 for a three-domain protein, wherein maxima in the CRs are observed when the equilibrium is poised such that one or a number of domains is unfolded in the most probable states in the absence of ligand. In fact, for all models a maximum in allosteric coupling is observed when one or a number of domains (or segments) are ID, indicating that the principles described here are not artifacts of the simplicity of the two-domain assumption and are extendable to multidomain systems.

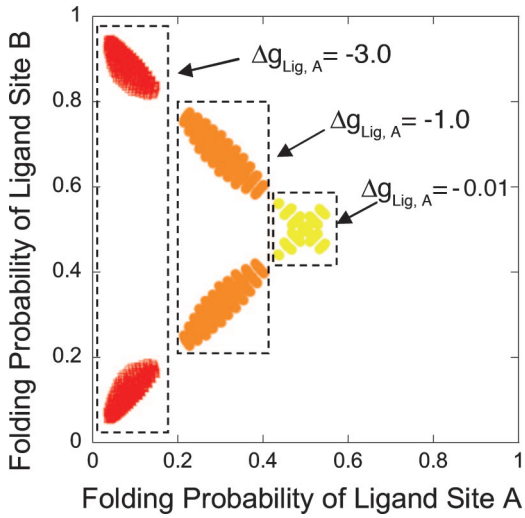


**Fig. 4.** The effect of binding energy on the optimum distribution for allosteric coupling. Plot of the folding probability of the ligand A site (i.e., $P_N + P_1$) vs. the folding probability of the ligand B site (i.e., $P_N + P_2$) showing only those parameter combinations where the CR exceeds a specified response threshold. Shown is the dependence of the optimum distribution on the binding free energy available to induce the allosteric transition. For binding energies of −0.01, −1.0, and −3.0 kcal/mol the corresponding values of $K_a[A]$ (from Eq. **4**) are 0.02, 4.4, and 157, respectively. For clarity, response thresholds of 0.23, 0.21, and 0.15 were used to show the maxima for $\Delta g_{Lig,A} = -0.01$, −1.0, and −3.0 kcal/mol, respectively.

**The Thermodynamic Basis for Coupling.** The results seen in Figs. 2 and 3 are at first glance somewhat surprising. Rather than being a continuous function of stability, optimum allosteric coupling is observed when the equilibrium is poised in either of two regions of thermodynamic parameter space (in the case of Fig. 2), when the ensemble is dominated by the state in which just domain I is unfolded (region 1 in Fig. 2), or when the ensemble is dominated by the state in which both domain I and II are unfolded (region 2 in Fig. 2). The origin of the bimodal response is that the two sites can be either positively or negatively coupled.

In the case of positive coupling (region 2), the effect of ligand is as described in Fig. 1*B*. Namely, the equilibrium is such that the unfolded state dominates the ensemble probabilities. Upon adding ligand A, those states with domain I folded will be preferentially stabilized. If domain I and II are positively coupled, then the energy of breaking the interaction between them will be positive (unfavorable), and states with only one domain unfolded will be highly improbable. As a result, stabilizing the binding site for ligand A will have the effect of also stabilizing the binding site for ligand B, simply because states where both domains are folded will be more probable.

In the case of negative coupling (region 1), the principles are the same but the effect is opposite. With negative coupling, the interaction energy, $\Delta g_{int}$, is negative, meaning that it is energetically unfavorable to have both domains folded at once, and states with one domain unfolded will dominate the ensemble probabilities. In this case, stabilization of domain I, via the binding of ligand A, will result in a destabilization of domain II, which will cause a decrease in affinity for ligand B. In either case, the results of this analysis are clear. Proteins containing ID domains are more able to propagate the effects of binding through different domains if the binding is coupled to the folding of the molecule.

Although the multimodal behavior in the distribution of CR values (Figs. 2 and 3) is caused by the existence of negative and positive coupling, it is important to note that the magnitude of the perturbation induced by ligand A (i.e., $\Delta g_{Lig,A}$) also plays a role in determining where the equilibrium is poised to elicit the maximum response. As shown in Fig. 4, in the limit where the system must respond to only minute changes in the fraction of molecules that are bound to A (e.g., $K_a[A] < 0.02$, which according to Eq. **4** gives $\Delta g_{Lig,A} \approx -0.01$ kcal/mol), the maximum response is obtained when each of the domains are unfolded 50% of the time. This scenario is likely to be rare, as allosteric effectors are usually bound tightly, and are molecules that have been selected by nature to act as effectors because they vary in concentration as a result of cellular or environmental changes. Nonetheless, even in cases when modest changes in energy are expected from ligand A (e.g., $K_a[A] = 4.4$, which according to Eq. **4** gives $\Delta g_{Lig,A} \approx -1.0$ kcal/mol), the equilibrium that will produce the optimum response involves a significant fraction of molecules wherein one or both domains are unfolded.

In any case, Fig. 4 indicates that where the equilibrium is poised before the addition of ligand A will depend on how much binding energy (Eq. **4**) is available to the system to elicit the desired signal. If the binding affinity for the effector ligand is low and/or the anticipated change in concentration of ligand is small relative to $K_a$, the ensemble will have a higher fraction of states that are structured. If, on the other hand, the binding affinity for the effector ligand is high and/or the anticipated change in effector concentration is large, the ensemble will be dominated by states that are partially or fully disordered. It is noteworthy that such a continuum in relative structure has been observed in disordered proteins (7).

**Relationship to Classical and Dynamic Allosteric Models.** Allostery has typically been discussed in terms of the Monod–Wyman–Changeux (MWC) (13) and Koshland–Nemethy–Filmer (KNF) (14) models, although both can be regarded as special cases of a more general allosteric model (15). The distinguishing aspects of these two models is that the MWC model describes equilibria between two macroscopic states, each of which can bind ligand (albeit, with different affinities), whereas the KNF model relies on an "induced-fit" mechanism, wherein binding is facilitated by only one form. In this respect, the current model more closely resembles the KNF formulation, although in the current formulation the binding incompetent states are disordered, as opposed to folded, compact structures. However, the current allosteric model differs from both the MWC and the KNF formulations in two important ways. First, unlike the MWC model, the ensemble model described here does not impose symmetry (i.e., high or low affinity for one ligand is not associated exclusively with high or low affinity for the second ligand). Second, unlike the KNF model, where the coupling energy is introduced as part of the binding energy for the ligand, the observed coupling between sites in the current ensemble model is a consequence of the intrinsic stabilities of the domains and the interactions between them. In other words, a single intrinsic binding constant describes the energy of interaction between the binding competent conformation of a domain and its putative ligand. Yet the observed binding affinity for each ligand, as well as the apparent coupling between the binding sites, is determined by the probability distribution that results from the conformational energies within the protein, as shown in Figs. 1 and 2. As a result, the current ensemble model provides a framework for investigating the underlying thermodynamic ground rules that relate regional stability with the observed allosteric coupling.

Of particular relevance to the current model are more recent studies that reveal the importance of conformational fluctuations around the canonical structure in mediating allosteric coupling (16–21). Because the classic allosteric models provide the mathematical relationships governing site-to-site coupling, but do not specify the structural basis of that coupling, fast local motions (18–20) can, in principle, be adequately captured, provided that differences in binding between microscopic elements of the same macroscopic state are much less than the average differences

between the different macroscopic states. This, however, is likely to be the case only when the conformational fluctuations are modest.

Interestingly, in the limit that the conformational fluctuations around the canonical structure resemble local folding/unfolding transitions (17, 21), the current model should provide considerable insight in interpreting experiments. Indeed, a number of examples have emerged, which suggest that conformational fluctuations around the canonical native structure may play an analogous role to intrinsic disorder in mediating allosteric coupling. First, conformational fluctuations within several single-domain proteins have been shown to be thermodynamically well represented as local order/disorder transitions, and these fluctuations often involve active-site residues (22–24). Second, site-to-site coupling has also been observed in a folded, conformationally heterogeneous protein in the absence of a connectivity pathway linking the coupled sites (25). Third, allosteric coupling energies were found to correlate with probes of the regional structural stability in a folded (and also conformationally heterogeneous) allosteric system (26). These observations, are difficult, if not impossible, to interpret in terms of a purely mechanical allosteric model. However, each is either a facet or a direct prediction of the ensemble representation, suggesting that the current model, despite its simplicity, may provide a useful framework for interpreting experiments, even for classic allosteric systems where fast local motions (18–20) are observed.

## Conclusions

The importance of the results presented here is twofold. First, they provide a general quantitative rationale for the observation that many regulatory proteins have ID regions; intrinsic disorder can maximize the ability to allosterically couple two sites. Second and equally as important, they can reveal the general thermodynamic ground rules for site-to-site coupling, wherein the ability to propagate the effects of binding are determined not necessarily by a mechanical pathway linking the two sites, but by the energetic balance within the protein (i.e., what states are most stable and what ligands can bind to each state). The significance of this result with regard to allosteric mechanisms cannot be overstated. As Fig. 2*A* reveals, the parameter combinations that produce optimal conditions for site-to-site coupling are highly degenerate, meaning that the stability of any one domain is not critical to the coupling. Changes in stability in one domain (or region) can be compensated by changes to another domain, or to changes in the interactions between domains. Indeed, the results indicate that the precise stabilization mechanism is not a determinant of the coupling at all. In effect, the sites can be coupled without the requirement of a specific network of interactions (or pathways) between the coupled sites. The relatively degenerate requirements for coupling described here appear to undermine the view that allostery is a precisely evolved property of proteins that relies on a specific mechanical pathway. In fact, these results suggest that site-to-site coupling in proteins can be robustly encoded (17, 21) and amenable to significant sequence divergence, a result that is also borne out in the low sequence similarities observed for the disordered regions within specific classes of proteins (9).

It is well established that proteins have modular structure (10), and that multidomain regulatory proteins often segregate the binding sites for each ligand into the different structural domains (9). A general mechanism, through which the different domains communicate, however, has proven elusive. The observation that ID domains and ID segments appear in disproportionately higher amounts in regulatory proteins (5) indicates that intrinsic disorder is important for their functional, regulatory role. The model described here provides a quantitative and unifying rationale for this observation and suggests that in evolving allosteric coupling between sites nature uses (at least in some cases) an ensemble-mediated mechanism. According to this mechanism, the stabilities of (and interactions between) the different domains in the protein produce an ensemble of states that is "optimally poised to respond" to binding. Upon binding, the ensemble is redistributed and the properties of the ensemble change accordingly. Because such a mechanism depends only on the relative stabilities of the domains and not the specific structural basis of that stability, it represents a radical departure from the classic mechanical view of allostery, wherein coupling would necessarily be facilitated through structural perturbations that extend from one binding site to the other.

Finally, several hypotheses have been put forward to explain why nature uses intrinsic disorder. These hypotheses, which include high-specificity/low-affinity binding (3, 5), rapid protein turnover (1), the ability to form large interaction surfaces (4, 5), and high specificity for multiple targets (1–5), focus primarily on how intrinsic disorder promotes molecular recognition. We do not challenge these benefits, nor do we suggest that intrinsic disorder is only used to facilitate long-range coupling. As described in ref. 8, intrinsic disorder is associated with a variety of functional roles not directly related to regulation. The current model, however, does provide significant insight into how a protein with ID domains can transmit signals from the binding of many different types of ligands, and it provides a quantitative framework to connect it to experimentally accessible quantities such as stability and binding affinity. In this respect, it is hoped that this model will broaden the scope of possible mechanisms and help guide in the interpretation and design of future experiments.

The ability to understand the determinants of allostery in proteins is the cornerstone to a quantitative description of biological processes. The model presented here represents a critical step in the development of a unifying framework that connects intrinsic disorder, conformational fluctuations, and classic models for allostery.

This work was supported by National Science Foundation Grant MCB-0446050, National Institutes of Health Grant GM-13747, and Welch Foundation Grant H-1461.

1. Wright PE, Dyson HJ (1999) *J Mol Biol* 293:321–331.
2. Uversky VN (2006) *Protein Sci* 11:739–756.
3. Uversky VN, Oldfield CJ, Dunker AK (2005) *J Mol Recognit* 18:343–384.
4. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CR, Hipps KW, *et al.* (2001) *J Mol Graphics Model* 19:26–59.
5. Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK (2006) *Biochemistry* 45:6873–6888.
6. Fink A (2005) *Curr Opin Struct Biol* 15:35–41.
7. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) *FEBS* 272:5129–5148.
8. Tompa P (2005) *FEBS Lett* 579:3346–3354.
9. Kumar R, Thompson EB (2005) *J Steroid Biochem Mol Biol* 94:383–394.
10. Brändén CI, Tooze J (1991) *An Introduction to Protein Structure* (Garland, New York).
11. Baldwin RL (1986) *Proc Natl Acad Sci USA* 83:8069–8072.
12. Lockless SW, Ranganathan R (1999) *Science* 286:295–299.
13. Monod J, Wyman J, Changeux P (1965) *J Mol Biol* 12:88–118.
14. Koshland JDE, Nemethy G, Filmer D (1966) *Biochemistry* 5:365–385.
15. Wyman J (1972) *Curr Top Cell Regul* 6:207–223.
16. Cooper A, Dryden DTF (1982) *Eur Biophys J* 11:103–109.
17. Pan H, Lee JC, Hilser VJ (2000) *Proc Natl Acad Sci USA* 97:12020–12025.
18. Fuentes EJ, Der CJ, Lee AL (2004) *J Mol Biol* 335:1105–1115.
19. Igumenova TI, Frederick KK, Wand AJ (2006) *Chem Rev* 106:1672–1699.
20. Popovych N, Sun S, Ebright RE, Kalodimos CG (2006) *Nat Struct Mol Biol* 13:831–838.
21. Liu T, Whitten ST, Hilser VJ (2006) *Protein Struct Funct Bioinform* 62:728–738.
22. Whitten ST, Garcia-Moreno EB, Hilser VJ (2005) *Proc Natl Acad Sci USA* 102:4282–4287.
23. Ferreon JC, Hamburger JB, Hilser VJ (2004) *J Am Chem Soc* 126:12774–12775.
24. Babu CR, Hilser VJ, Wand AJ (2004) *Nat Struct Mol Biol* 11:352–357.
25. Clarkson MW, Gilmore SA, Edgell MH, Lee AL (2006) *Biochemistry* 45:7693–7699.
26. Gekko K, Obu N, Li J, Lee JC (2004) *Biochemistry* 43:3844–3852.