

Can allosteric regulation be predicted from structure?

Ernesto Freire*

Department of Biology, The Johns Hopkins University, Baltimore, MD 21218

During the last decade, the classical view of a protein existing in equilibrium between discrete conformational states has undergone a radical transformation. Experimental observations, particularly those obtained from NMR-detected hydrogen/deuterium exchange, indicate that even under native conditions proteins need to be considered as complex statistical ensembles (see for example refs. 1–13). In the view pioneered by Englander and coworkers (1, 2, 4, 7, 8, 14) proteins undergo local unfolding reactions scattered throughout their entire structures. These unfolding reactions occur independently of each other, may involve only a few amino acids, and give rise to a large number of states in which each state is defined by the presence of one or several locally unfolded regions. This collection of states defines the native-state ensemble. The immediate, and perhaps most relevant, consequence of these observations is that the Gibbs energy of stabilization of a protein is not uniformly distributed throughout its three-dimensional structure. There are regions with high stability constants and regions with low stability constants (13, 15). Furthermore, under native conditions cooperativity appears to be local or regional rather than global. The implications of these findings have been mostly discussed in relation to the protein folding problem; however, their biological implications are tremendous because the native state is the functionally relevant state. If cooperativity is local or regional rather than global, how do distal sites communicate with each other? How do regulatory allosteric interactions take place? How are the functional properties of a protein related to the distribution of states within the native ensemble, and how does ligand binding affect this distribution and modulate function? Is the location of stable and unstable regions a random event or it is dictated by functional considerations? How are protein function and stability related?

The tools to begin addressing those issues in a systematic way became available with the observation that a structure-based algorithm (COREX) was able to model reasonably well the hydrogen exchange protection data obtained for many proteins (13, 15–19). Within the frame-

work of the COREX algorithm the protein is considered as a statistical ensemble in which each state is characterized by having some region or regions in a non-folded state. These regions can be as small as four or five amino acids or as large as the entire protein. Depending on the size of the protein the computation is performed either exhaustively or by using a sampling technique (19). The Gibbs energy of all of the resulting states and their respective probabilities are calculated in terms of an empirical structural parameterization of the Gibbs energy (summarized in ref. 20). What the COREX algorithm produces is a list of possible conformations and their respective probabilities, i.e., a probability distribution function. As such it can be used to examine the effects of ligands and other chemical or physical factors on that distribution. Previously (21) it was shown that the incorporation of the ligand linkage equations into the COREX algorithm (CORE_BIND) correctly predicted the propagation of binding effects through the structure of hen egg white lysozyme upon binding of a specific antibody. The paper by Pan *et al.* (22) in this issue of PNAS explores the linkage between binding sites in *Escherichia coli* dihydrofolate reductase and makes a convincing case that the coupling between these sites is also mediated by shifts in the probability distribution of the conformational ensemble.

The COREX algorithm produces a snapshot of the distribution of states existing under equilibrium conditions. For an ergodic system this distribution is identical to the one that would be obtained if a single protein molecule were observed over a period sufficiently long for thermodynamic averaging. Accordingly, ensemble properties can be mapped into individual molecules. This is illustrated in Fig. 1 for the structural stability of individual residues in a Src homology 3 molecule (18). In the ensemble view, the stability of an individual residue is proportional to the summed probabilities of all of the states in which that residue is in the native state. For an individual molecule, the structural stability of a residue is proportional to the fractional amount of time that the residue spends in the native-state conformation. A highly stable residue will spend most of

the time in the native state, whereas a relatively unstable residue will make excursions into different regions of conformational space. These excursions may arise from different processes involving local as well as longer-range fluctuations. For an ergodic system, ensemble and time averages are equivalent.

As pointed out by Pan *et al.* (22) classical linkage theories do not address the mechanism by which different binding sites communicate. The ensemble view created by the COREX algorithm provides an opportunity to examine the long-range effects of ligand binding and allosteric regulation. We can imagine that different conformational states might have different functional properties and that, consequently, a redistribution in the population of states may result in functional changes. A change in the distribution of states is triggered by changes in the Gibbs energy of the states that define the ensemble. States with lower Gibbs energies will be preferentially populated with respect to states with higher Gibbs energies. Therefore, the expression of a specific functional property can be triggered by a decrease in the Gibbs energy of those states that exhibit that property. In biological systems, this is usually accomplished by ligand molecules (activators or inhibitors) that selectively bind to those states that possess the selected property. In the presence of a ligand *X*, the Gibbs energy of any arbitrary state within the ensemble will be affected by an amount that depends on its binding affinity for the ligand:

$$\Delta G_i = \Delta G_i^0 - RT \ln \frac{(1 + K_{a,i}[X])}{(1 + K_{a,0}[X])}, \quad [1]$$

where ΔG_i^0 is the Gibbs energy of state *i* in the absence of the ligand, $K_{a,0}$ is the binding constant to the reference state, and $K_{a,i}$ the binding constant to state *i*. Those states that are able to bind the ligand will be stabilized with respect to those states that are not able to bind the ligand, causing a change in the probability distribution of states. Fig. 2 illustrates this

See companion paper on page 12020.

*E-mail: ef@jhu.edu.

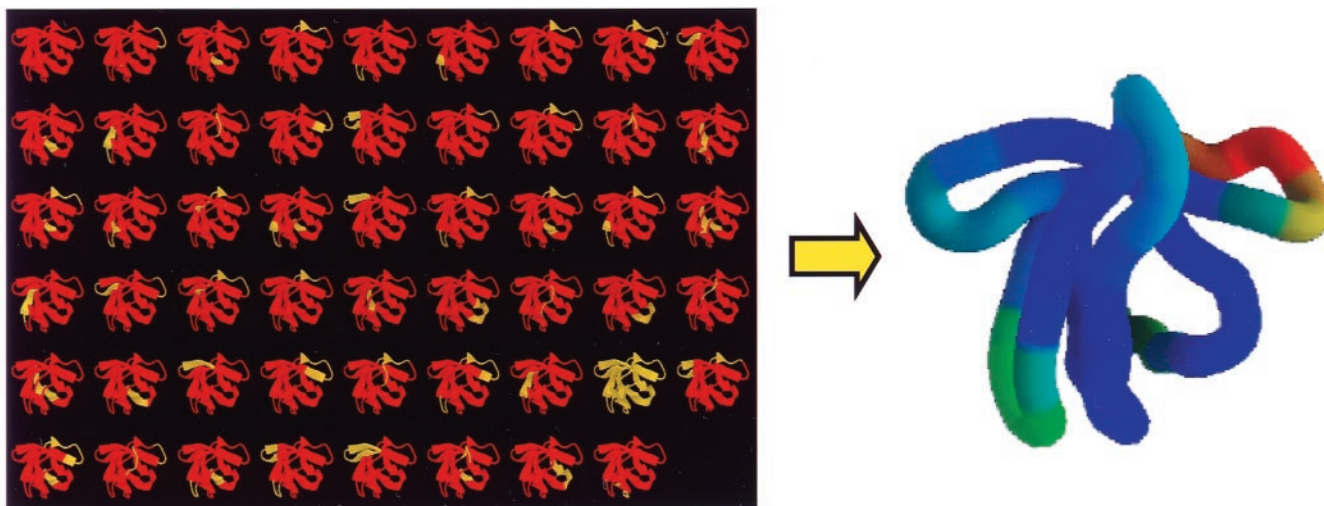


Fig. 1. The COREX algorithm produces an ensemble representation of a protein. This representation can be mapped into a single molecule, because for an ergodic system ensemble and time averages are equivalent.

situation with the binding of a proline-rich helical peptide to a Src homology 3 domain. The protein ensemble can be considered as being composed of states in which the binding site is intact and therefore binding-competent and states in which the binding site is not formed and therefore unable to bind the ligand. The states that are binding-competent will be stabilized according to Eq. 1, resulting in a shift in the probability distribution. The residue stabilities calculated in the presence of the ligand can be mapped into an individual protein molecule as well as the differences in Gibbs energy per residue, as shown in Fig. 2. A notable result is that not all amino acids are affected equally, a consequence of the lack of global coop-

erativity in the native state. Under those conditions the transmission of binding signals to distal sites involves only a subset of residues. Their identification is important not only from the point of understanding a fundamental biological phenomena but also from the point of view of protein engineering and drug design.

Communication between sites occurs when their behavior is coupled. For regulatory and active binding sites this coupling is expressed in a correlation between their binding competencies. For allosteric inhibition, binding to the regulatory site must trigger a destabilization of the active conformation of the catalytic site such that substrate either does not bind or binds in a nonproductive fashion. In this

situation, the native-state ensemble can be considered as being primarily populated by two distinct subensembles: one subensemble of conformations that are catalytically active but are unable to bind the inhibitor, and one subensemble of molecules that lack biological activity but are able to bind the inhibitor. The presence of the inhibitor will shift the equilibrium toward the inactive conformations because those conformations are the ones

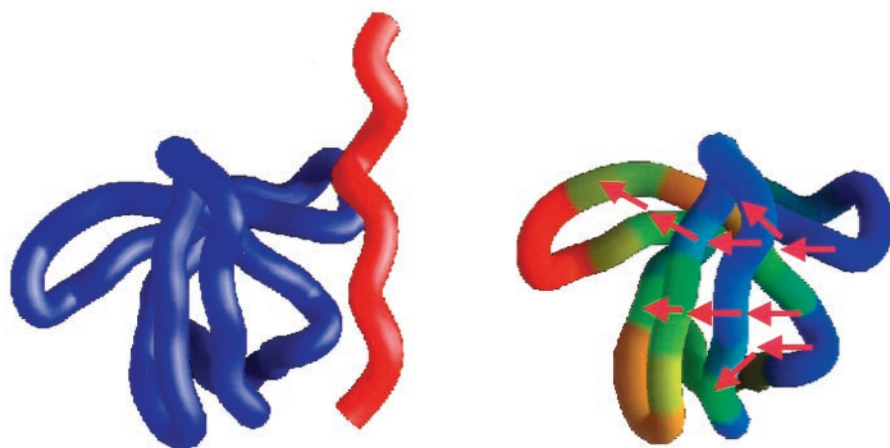


Fig. 2. The binding of a proline-containing peptide (red) to the Src homology 3 domain (blue). Ligand binding stabilizes only those conformations that are binding-competent, inducing a redistribution in the conformational ensemble. The effects of binding can be mapped into a single molecule as shown on the right. The blue regions are the ones most affected by binding and the red regions the ones least affected. Notably, not all amino acids are predicted to be equally affected because of the lack of global cooperativity. A consequence of this result is that communication between distal sites involves only a subset of residues as elegantly shown by Pan *et al.* (22) for dihydrofolate reductase.

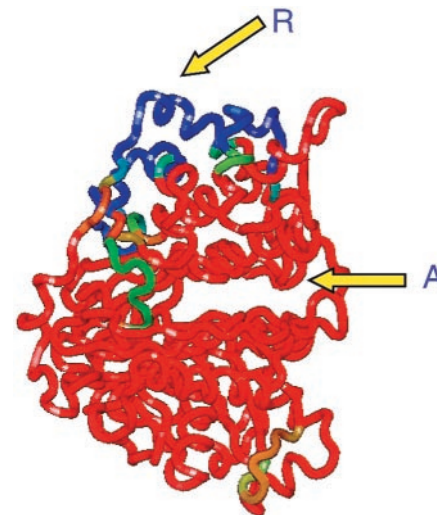


Fig. 3. Results of COREX and CORE_BIND analysis of glycerol kinase and its allosteric inhibitor IIAGlc (19). The binding site for IIAGlc (R) to glycerol kinase contains regions of low structural stability. In the absence of IIAGlc, those regions do not interact strongly with the rest of the protein. Upon binding IIAGlc, those regions become structurally stable and initiate a sequence of interactions that shift the catalytic site into an inactive conformation. According to the computational results only a well-defined set of residues participates in the allosteric coupling of both sites.

that are able to bind the inhibitor. Again this redistribution of the ensemble can be mapped into a single molecule to identify those residues that are implicated in the communication between sites. The allosteric enzyme glycerol kinase provides a clear illustration of the above situation as illustrated in Fig. 3. An allosteric inhibitor of this enzyme is the protein IIAGlc. IIAGlc binds to glycerol kinase at a site

located more than 30 Å away from the active site. The binding of IIAGlc stabilizes this region into a conformation capable of establishing specific contacts with adjacent regions that result into the stabilization of the inactive conformation of the catalytic site. According to the results of the COREX and CORE_BIND analysis only a small set of residues are involved in the allosteric coupling between

regulatory and catalytic sites. The results obtained for dihydrofolate reductase by Pan *et al.* (22) emphasize those conclusions and provide new tools for the investigation of long-range communication in proteins.

This work was supported by grants from the National Institutes of Health (GM 51362 and GM 57144) and the National Science Foundation (MCB-9816661).

1. Jeng, M.-F., Englander, S. W., Elöve, G. A., Wand, A. J. & Roder, H. (1990) *Biochemistry* **29**, 10433–10437.
2. Jeng, M.-F. & Englander, S. W. (1991) *J. Mol. Biol.* **221**, 1045–1061.
3. Radford, S. E., Buck, M., Topping, K. D., Dobson, C. M. & Evans, P. A. (1992) *Proteins* **14**, 237–248.
4. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993) *Proteins* **17**, 75–86.
5. Kim, K.-S. & Woodward, C. (1993) *Biochemistry* **32**, 9609–9613.
6. Woodward, C. (1993) *Trends Biochem. Sci.* **18**, 359–360.
7. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995) *Science* **269**, 192–197.
8. Milne, J. S., Mayne, L., Roder, H., Wand, A. J. & Englander, S. W. (1998) *Protein Sci.* **7**, 739–745.
9. Schulman, B. A., Redfield, C., Peng, Z., Dobson, C. M. & Kim, P. S. (1995) *J. Mol. Biol.* **253**, 651–657.
10. Chamberlain, A. K., Handel, T. M. & Marqusee, S. (1996) *Nat. Struct. Biol.* **3**, 782–787.
11. Swint-Kruse, L. & Robertson, A. D. (1996) *Biochemistry* **35**, 171–180.
12. Dabora, J. M. & Marqusee, S. (1994) *Protein Sci.* **3**, 1401–1408.
13. Hilser, V. J., Oas, T., Dowdy, D. & Freire, E. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 9903–9908.
14. Milne, J. S., Xu, Y., Mayne, L. C. & Englander, S. W. (1999) *J. Mol. Biol.* **290**, 811–822.
15. Hilser, V. J. & Freire, E. (1996) *J. Mol. Biol.* **262**, 756–772.
16. Hilser, V. J., Townsend, B. D. & Freire, E. (1997) *Biophys. Chem.* **64**, 69–79.
17. Hilser, V. J. & Freire, E. (1997) *Proteins* **27**, 171–183.
18. Sadqi, M., Casares, S., Abril, M. A., Mayorga, O. L., Conejero-Lara, F. & Freire, E. (1999) *Biochemistry* **38**, 8899–8906.
19. Luque, I. & Freire, E. (2000) *Proteins* **4**, 63–71.
20. Luque, I. & Freire, E. (1998) *Methods Enzymol.* **295**, 100–127.
21. Freire, E. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 10118–10122.
22. Pan, H., Lee, J. C. & Hilser, V. J. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 12020–12025.