Engineering and design of ligand-induced conformational change in proteins
Laura S Mizoue* and Walter J C hazin†

The ability to manipulate ligand-induced conformational change, although representing a major challenge to the protein engineer, is an essential end point in efforts to produce novel functional proteins for biotechnology and therapeutic applications. Progress towards this goal requires determining not only what factors control the fold and stability of a protein, but also how ligand binding alters the complex conformational/energetic landscape. Important strides are being made on several fronts, including understanding the origin of long-range effects and allosteric structural mechanisms, using both experimental and theoretical approaches.

Addresses
Departments of Biochemistry and Physics, Center for Structural Biology, 896 PRB, Vanderbilt University, Nashville, TN 37232-0146, USA
* e-mail: l.mizoue@vanderbilt.edu
† e-mail: walter.chazin@vanderbilt.edu

0959-440X/02/$ — see front matter © 2002 Elsevier Science Ltd. All rights reserved.

Abbreviations
CaM calmodulin
Clb9 calbindin D9k
MBP maltose-binding protein
SH Src homology

Introduction
Significant advances in protein engineering and design over the past ten years have prompted the field to expand in a variety of exciting new directions geared towards generating the ability to specify function. Intermolecular interaction lies at the core of protein function in a wide range of fundamental biochemical processes. In many instances, binding induces changes in structure and dynamics at a remote site (allostery) that allow, for example, signal transduction across large distances. In this review, we focus on recent advances relating to ligand-induced conformational changes in proteins.

The first three sections of the review cover four experimental design efforts that have focused on controlling protein conformation. In each case, the design template is a protein known to undergo substantial conformational changes upon ligand binding and the objective is to mutate residues so that one state is preferentially stabilized over the other (Figure 1). These studies represent very early steps toward the ultimate goal of specifying ligand-induced conformational changes. To achieve this larger objective, it will be necessary to incorporate increasingly sophisticated tools that will require a more detailed understanding of the relationships between protein sequence, structure, dynamics and function. Therefore, the final two sections of this review focus on recent reports that we feel provide important insights and information that will aid future design efforts.

Conformational control in EF-hand proteins
Extensive structural and biophysical studies of the EF-hand family of Ca2+-binding proteins have revealed two primary functional classes [1,2]. Ca2+ sensors, such as calmodulin (CaM), transduce Ca2+ signals, whereas Ca2+ signal modulators, such as calbindin D9k (Clb9), modulate the shape and/or duration of Ca2+ signals or help maintain Ca2+ homeostasis. Although CaM and Clb9 are highly homologous (~25% sequence identity) and have very similar apo state structures, they display striking differences in their conformational response to Ca2+ binding (Figure 1a,b). Each of the two EF-hand domains of CaM undergoes a pronounced Ca2+-induced change into an open conformation that is characterized by an exposed hydrophobic patch [3–5]. Clb9, on the other hand, remains in a closed conformation upon Ca2+ binding that is similar (but not identical) to its apo state [6,7].

Recent efforts from two laboratories have been directed at determining why there is such a drastic difference in the Ca2+-induced changes of the two proteins. These groups developed very similar approaches to the problem, in one case using the CaM framework and in the other using the Clb9 framework. Both groups developed design hypotheses based on comparative analysis of the sequences of the two proteins, of the apo and Ca2+-loaded structures of CaM and Clb9, and of a homology model of Clb9 in the open conformation.

Ababou and Desjarlais [8*] used the structural comparisons to estimate the contributions of different residues to the solvation energetics of the open and closed conformations. They noted that polar amino acids (glutamine and lysine) are present at positions 41 and 75 of CaM, whereas non-polar amino acids (leucine and isoleucine) occupy the structurally equivalent positions in Clb9. These residues are partially buried in the closed conformation of CaM, but become fully solvent exposed upon Ca2+ binding. They designed Gln41→Leu and Lys75→Ile mutations of CaM, anticipating that hydrophobic residues at these positions would be readily accommodated in the closed conformation, but would be destabilizing in the open conformation. These mutations led to a significant increase in the stability of the apo protein, a decrease in the Ca2+ affinity and a reduction in the ability to bind ANS (a small-molecule fluorescent probe for exposed hydrophobic surfaces), consistent with a more favorable closed conformation and/or a less favorable open conformation.
It is important to note that Gln41 and Lys75 are located far from the Ca$^{2+}$-binding sites, yet their effects on Ca$^{2+}$ affinity are significant. Leu39 is close in space to these residues and is structurally homologous to Phe36 in Clb9. Phenylalanine is rarely seen at this position in CaM-like domains. Moreover, in the homology model of Clb9 in the open conformation, Phe36 is highly solvent exposed. Consequently, the authors introduced a Leu39→Phe mutation in the background of the wild-type and mutant proteins from the previous study [9]. Interestingly, although the Ca$^{2+}$ affinities and the relative stabilities of the closed and open conformations were altered by this mutation, the effects were not fully additive in all cases.

In a highly complementary study, Nelson, Chazin and co-workers examined the effect of a Phe36→Gly mutation in Clb9. Their ultimate design objective is to re-engineer the biochemical function of the signal modulator Clb9 so that it responds to Ca$^{2+}$ binding in the manner of the Ca$^{2+}$ sensor CaM [10]. Using very detailed structural comparisons [11], they identified Phe36 as an important determinant of the packing between helices I and II. They anticipated that substituting a glycine at this position would change this interface in the apo protein, making it more like the closed conformation of CaM. The mutation was found to have little effect on Ca$^{2+}$ affinity, on the stability of the apo protein or on the structure of the Ca$^{2+}$-loaded state [12•]. However, the 3D structure was determined in the apo state and showed that the helix I–helix II interface was clearly altered, demonstrating the success of the first step towards functional re-engineering of the protein.

**MBP-based biosensors: modulating ligand binding by controlling conformation**

*Escherichia coli* maltose-binding protein (MBP) is a monomer composed of two globular domains connected by a hinge region. The binding site is located at the interface of the two domains and the protein undergoes a large conformational change upon complexation [13,14]. This ligand-induced switch involves a reorientation of the two domains from an open (apo) to a closed (bound) state through bending and twisting of the hinge region (Figure 1c).

By engineering simple point mutations that alter the stability of the open and closed conformers, Marvin and Hellinga [15••] have shown that they can ‘tune’ the binding affinity of MBP. To identify potential sites for mutation, they first compared the crystal structures of the apo and ligand-bound proteins, looking for regions that are different in the two states but remote from the ligand binding pocket [16]. Ile329 lies in a crevice of the hinge region and is buried in the open state. Based on simple steric arguments, they predicted that mutating Ile329 to a residue with a bulkier sidechain would destabilize the open conformation. Replacement with phenylalanine, tyrosine and tryptophan led to increased binding affinities that roughly correlate with the increase in sidechain...
volume, whereas replacement with a smaller alanine sidechain decreased the binding affinity. These results are consistent with the implication that destabilization of the open conformation leads to increased binding affinity. They tested another site, Ala96, where a larger sidechain was predicted to stabilize the closed conformation. The introduction of an Ala96→Trp mutation into the Ile329→Trp background showed 3-fold and 60-fold increases in binding affinity relative to the single site mutant and wild-type protein, respectively. This strategy has been incorporated into iterative designs of MBP-based metallosensors and metalloenzymes [17,18].

**Integrin I domains: conformations made to order**

Integrins are heterodimeric transmembrane receptors that mediate cell adhesion. The I domain of several integrin α subunits plays a major role in binding extracellular ligands; conformational changes in this domain are believed to trigger integrin activation. Crystal structures indicate that I domains switch from a closed to an open conformation upon ligand binding (Figure 1d). The open form is believed to correspond to the active state of the receptor [19–24].

Springer, Mayo and colleagues [25•] have taken a computational approach to design hydrophobic core variants of the Mac1 integrin I domain that preferentially adopt either an open or a closed conformation. Their computer algorithm [26] identifies optimal amino acid sequences for a given target structure. Four of the designed sequences were expressed in the context of the intact heterodimeric receptor and assayed for ligand binding. The open conformer designs contained 8–13 mutations and showed increased binding affinity relative to the wild type. In contrast, the closed conformer design, which contained four mutations, showed slightly decreased binding affinity. Furthermore, when fused to an artificial transmembrane domain and expressed in isolation, the open designs still displayed high affinity for ligand, whereas the closed design and the wild-type domain showed no appreciable binding.

**The transmission of long-range effects: key residues and coupling pathways**

When a protein binds ligand, the binding effects are transmitted through the structure. Although it is generally accepted that interaction networks exist within structural elements and globular domains, the mechanistic basis for these effects remains a topic of debate. It is not yet possible to generally predict these connectivity pathways given a protein structure.

Lockless and Ranganathan [27••] have taken an interesting approach to identify energetically coupled residues that uses evolutionary data from multiple sequence alignments of large protein families. The basic assumption is that a functional linkage between two positions will mutually constrain evolution at these sites and this will be reflected in the statistical coupling of their amino acid distributions. They tested their theory on the PDZ domain family and found that position 76, a site known to affect ligand specificity, was statistically coupled to only a small subset of residues. Some of the coupled positions were spatially distant from position 76, but were connected together via van der Waals contacts. These results suggest the occurrence of interaction pathways that enable transmission of local perturbations and provide the basis for allosteric effects.

The studies of the Clb9 mutant discussed previously appear to support the concept that key residues control energy conduction within proteins. The structural effects of the Phe36→Gly point mutation were quite substantial and extended not only into the adjacent helical interface, as intended, but also to regions far from the mutation. In fact, a hypothetical network of structural transmission was proposed based on comparisons of the mutant and wild-type apo state structures [12•].

Hilser and co-workers [28•] used a different strategy to identify residues involved in allosteric coupling between regulatory and catalytic sites of E. coli dihydrofolate reductase. In this approach, a structure-based thermodynamic algorithm was used to explore in silico how mutational or binding effects propagate from their point of origin [29–31]. One of the important results of this study was their discovery of energetically coupled residues that showed no obvious structural connection to the binding site. This implies that communication between sites does not necessarily occur through a series of conformational distortions. Instead, they argue that their observations represent a ligand-induced redistribution of the conformational ensemble.

**Population distributions of conformational states**

An increasingly popular theory views ligand binding as a shift in the energy landscape (reviewed in [32,33•]). This stems from the fact that proteins do not exist as static structures but as dynamic ensembles. All energetically accessible conformations are part of the equilibrium ensemble, even if the population of a given state is effectively nil under a certain set of conditions. The conformers with the most favorable sites will bind ligand and be preferentially stabilized.

Recent NMR spectroscopy studies have provided experimental support for this paradigm [34,35]. Of particular relevance here is a study of NtrC, a bacterial ‘two-component system’ signaling protein [36]. NMR relaxation experiments were used to demonstrate that the protein exists as an equilibrium population of inactive and active conformations, and that phosphorylation shifts the equilibrium towards the active species. Similarly, the role of Ca**2**+ binding in EF-hand proteins can be viewed as a ligand-induced shift in the equilibrium between open and closed conformations [37–39].
It is becoming increasingly clear that a static view of protein structure is insufficient to understand the energetic effects of ligand binding (e.g. [40]). For example, NMR relaxation and hydrogen-deuteron exchange measurements of the c-Src SH3 domain have shown that the dynamic response to ligand binding can extend to regions remote from the binding interface [41]. Freire and co-workers [42,43] have observed that low structural stability is a particularly important characteristic of regulatory sites. This is supported by a hydroxy radical footprinting study of BirA, the E. coli repressor of biotin synthesis, that highlights the significance of flexible surface loops in the transmission of allosteric information [44].

Conclusions

In natural systems, the propagation of conformational changes has been optimized by evolution so that thermodynamic and kinetic properties are perfectly matched to the requirements of function [45]. The protein engineer is making very good progress in controlling protein stability and fold. However, engineering a delicately tuned energy balance will be much more complicated. Clear progress has been made on both experimental and theoretical fronts towards understanding the factors that control the conformational changes induced by ligand binding and these results are encouraging. These and other approaches, such as directed random mutagenesis combined with intelligent screening, will all be required to develop the tools needed to achieve the ultimate goal of producing novel functional proteins.

Acknowledgements

Work on protein engineering and design in our laboratory is supported by operating grants from the National Institutes of Health (GM40120 and GM62112). We thank Shibani Bhattacharya and Jonathan Shechman for critical reading of the manuscript.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

* of special interest
** of outstanding interest


8. Ababou A, Desjarlais JR: Solvation energetics and conformational change in EF-hand proteins. Protein Sci 2001, 10:301-312. The replacement of polar with nonpolar residues at two positions in the N-terminal domain of CaM leads to increased stability of the apo state and a decrease in the Ca2+ affinity. The authors propose that the burial of polar groups is an important factor contributing to Ca2+-induced conformational change in EF-hand proteins.


The structure of a Clb9 point mutant demonstrates the success of a design to alter the packing interactions of a specific helical interface only in the apo protein. The authors propose a pathway for structural coupling through the protein to explain the observation of additional long-range structural perturbations.


An excellent example of rationally designed mutations that stabilize one of two conformational states. By introducing mutations at remote sites that affect the conformational equilibrium between the two states, the authors are able to tune the binding affinity of MBP.


25. Shimakawa M, Shifman JM, Jing H, Takagi J, Mayo SL, Springer TA: Computational design of an integrin I domain stabilized in the open high affinity conformation. Nat Struct Biol 2000, 7:674-678. Amino acid sequences designed to preferentially adopt either the closed or the open conformation of the integrin I domain were selected using a computer algorithm. The designs based on the open state structure bind ligand with higher affinities than either the wild-type protein or the designs based on the closed state structure.
The redistribution of the ensemble of microstates. The authors assert that the energetic couplings are mediated by allosteric communication between binding sites in POZ domain families, and the prediction of coupled sites agrees well with those identified from mutational studies.

An ensemble-based model was used to investigate the mechanism of binding in terms of shifts in the populations of conformational substates sampled by a protein.


The propagation of binding interactions to remote sites in proteins: analysis of the binding of the monoclonal antibody D1.3 to lysozyme. Proc Natl Acad Sci USA 1997, 94:10118-10122.

Luque I, Freire E: Structural stability of binding sites: consequences for binding affinity and allosteric effects. Prots 2000, 4:43-71. A structure-based thermodynamic analysis of 16 different proteins reveals that their binding sites contain regions of both low and high structural stability. In allosteric enzymes, the low stability regions of regulatory sites appear to be critical for transducing signals to the catalytic site.

