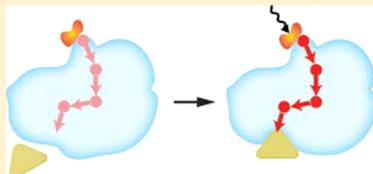


Controlling Allosteric Networks in Proteins

Nikolay V. Dokholyan*

Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, North Carolina 27599, United States

ABSTRACT: Allosteric transition, defined as conformational changes induced by ligand binding, is one of the fundamental properties of proteins. Allostery has been observed and characterized in many proteins, and has been recently utilized to control protein function via regulation of protein activity. Here, we review the physical and evolutionary origin of protein allostery, as well as its importance to protein regulation, drug discovery, and biological processes in living systems. We describe recently developed approaches to identify allosteric pathways, connected sets of pairwise interactions that are responsible for propagation of conformational change from the ligand-binding site to a distal functional site. We then present experimental and computational protein engineering approaches for control of protein function by modulation of allosteric sites. As an example of application of these approaches, we describe a synergistic computational and experimental approach to rescue the cystic-fibrosis-associated protein cystic fibrosis transmembrane conductance regulator, which upon deletion of a single residue misfolds and causes disease. This example demonstrates the power of allosteric manipulation in proteins to both elucidate mechanisms of molecular function and to develop therapeutic strategies that rescue those functions. Allosteric control of proteins provides a tool to shine a light on the complex cascades of cellular processes and facilitate unprecedented interrogation of biological systems.



CONTENTS

1. Introduction	6463
1.1. Allostery	6463
1.2. Allosteric Regulation	6465
2. Physical Origin of Allostery	6466
2.1. Allostery from a Graph Theory Perspective	6466
2.2. Evolution of Allostery	6468
2.3. Identifying Allosteric Pathways	6468
2.3.1. Physics-Based Approach	6468
2.3.2. Evolutionary Approach	6471
2.3.3. Nuclear Magnetic Resonance	6472
2.3.4. ϕ -Value Analysis	6473
2.3.5. Double Mutant Cycles	6473
2.4. Identifying Allosteric Sites Using Small Molecules Screening and Crystallography	6474
3. Control of Allostery Using Ligands	6474
3.1. Protein Activity Regulation through Mutually Exclusive Folding	6474
3.2. Molecular Switch Design by in Vitro Recombination of Nonhomologous Genes	6475
3.3. Protein Activity Regulation Using Small Molecule Ligands	6475
3.4. Inducing Large Conformational Changes Using Metals	6477
3.5. Regulation through Engineered Allostery	6477
3.5.1. Grafting Allostery from One Orthologous Protein to Another	6477
3.5.2. Engineering Allosteric Regulation Using Next-Generation Sequencing	6478
4. Allosteric Control of Proteins with Light	6478
5. Controlling Protein Dynamics and Stability by Targeting Allosteric Networks	6478
5.1. Allosteric Rescue of CFTR	6478

6. Outlook	6479
Author Information	6479
Corresponding Author	6479
Notes	6479
Biography	6479
Acknowledgments	6479
References	6479

1. INTRODUCTION

1.1. Allostery

The original definition of *allosteric transition*¹ is based on conformational change, or *induced fit*,² of the active site of a target protein upon binding of an effector to this protein at a specific site (*allosteric site*), distinct from the active site (Figure 1). This allostery phenomenon has been reported for a wealth of biological molecules;^{3–15} some of the first proteins discovered to feature allosteric response to a ligand binding were hemoglobin,¹⁶ L-threonine deaminase,¹ aspartic-transcarbamylase,¹⁷ aspartokinases I and II,¹⁸ and homoserine-dehydrogenase.¹⁹ The original view of allostery was structure-centric, whereby the tense state *T* converts into the relaxed state *R* upon ligand binding. Monod, Wyman, and Changeux²⁰ proposed a simple statistical mechanics (MWC) model of allostery that was able to recapitulate allosteric regulation of protein activity by ligands for a plethora of proteins, making this model an iconic model for years to come.

Special Issue: Protein Ensembles and Allostery

Received: September 15, 2015

Published: February 19, 2016

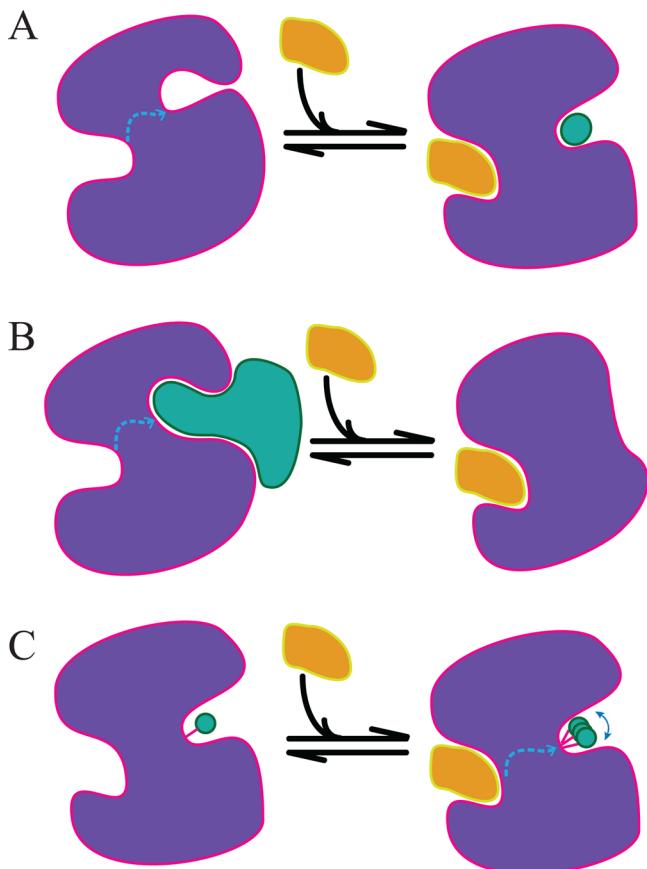


Figure 1. Examples of allostery. (A) Ligand-driven ligand binding: the association of one ligand (yellow) induces conformational change in a protein (purple) that makes it competent to bind another ligand (green). (B) Ligand-driven disruption of protein–protein interactions: the association of a ligand induces conformational change in a protein that abolishes interaction with a partner protein. (C) Ligand-driven dynamic coupling: the association of a ligand induces changes in the dynamics of the active site (green) thereby affecting its activity.

Over the years, the original notion of allostery was extended, especially by the revolutionary work of Cooper and Dryden,²¹ who proposed a dynamics based model of allosteric. In their model, states R and T are structurally similar, if not indistinguishable, and the difference between the states appears in the fluctuations of the active site upon effector molecule binding. The model of Cooper and Dryden was particularly important as it departed from the purely structural view to an ensemble view of protein allosteric. This more physical approach was based on prior observations made by many theoretical,²² experimental,^{23,24} and computational^{25–27} studies suggesting the importance of protein conformational dynamics to functioning.

Due to recent interest in controlling proteins using allosteric sites, many scientists have focused on the ensemble description of allosteric.^{9,28–39} Cui and Karplus argued that the “new view” of allosteric, emphasizing “population shifts,” is, in fact, an “old view”.⁴⁰ Successful examples of engineered allosterically controlled protein activity showed critical dependence on ensemble description of allosteric. Hence, in a broader sense, allosteric is the phenomenon in which two sites on a single biological molecule are dynamically coupled despite being outside of direct physical interaction range. This coupling is biased by ligand binding to the allosteric sites. The power of the ensemble description of protein allosteric is that it shifts the notion of two deterministic states to probabilistic description of ensembles describing these states. In the limit when the ensemble population of each state is equal to one, both definitions, structural and ensemble, exactly coincide. It is also possible to relate ensemble to structural definitions if one corrects by the probabilities of being in the corresponding states T and R .

Proteins are not the only biological molecules that feature allosteric transitions. Functions of DNA⁴¹ and many RNA molecules heavily rely on allosteric regulation. Perhaps one of the most striking examples is riboswitches (**Figure 2A**), portions of mRNA (mRNA) molecules that undergo conformational changes upon binding of metabolites, thereby inhibiting or initiating protein expression by this mRNA. Many riboswitches add self-regulating abilities to mRNA in response

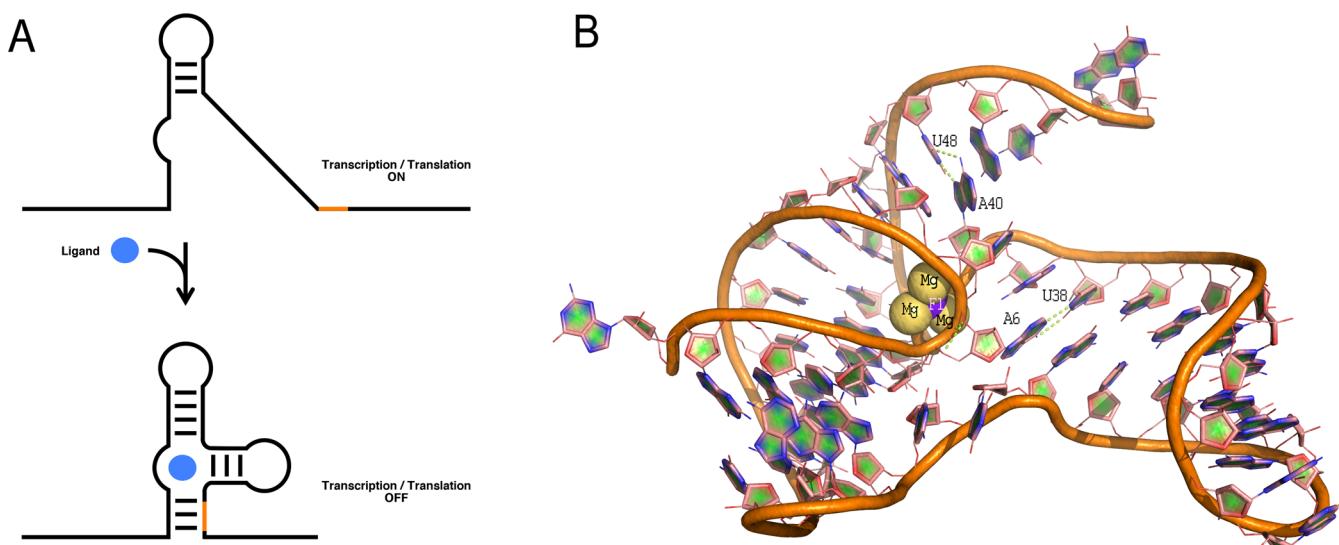


Figure 2. Riboswitches. (A) Mechanism of action of a riboswitch: upon ligand binding, the transcription/translation of downstream mRNA halts. (B) Three-dimensional structure of the fluoride riboswitch from *T. petrophila* determined by X-ray crystallography.⁴⁵

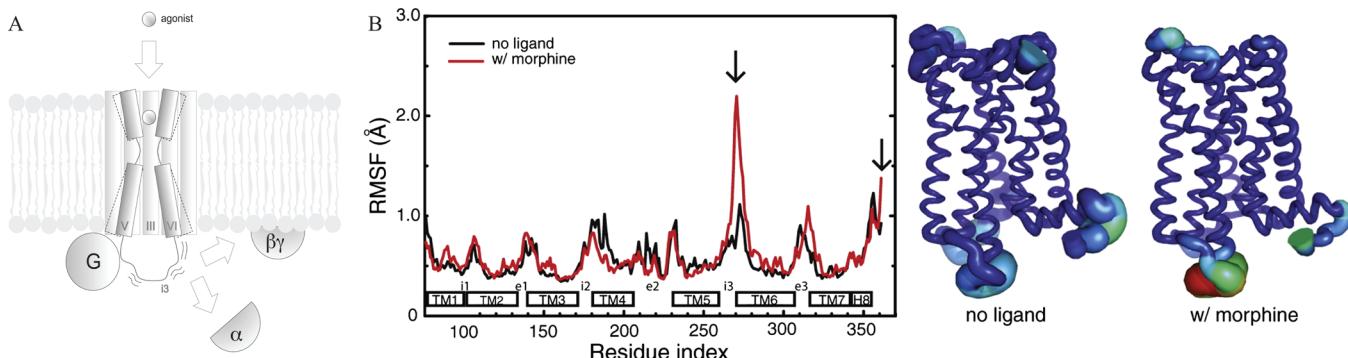


Figure 3. Allostery in GPCR. (A) Mechanism of GPCR activation: upon ligand binding on the extracellular side of the protein, GPCR undergoes conformational changes that result in altering dynamics on the cytoplasmic side, particularly i3 loop, which results in dissociation of G protein complex from GPCR and activation of downstream pathways. (B) An example of altered dynamics described in (A) was recently demonstrated in the case of μ -opioid receptor activation by Serohijos et al.³³¹ The root-mean-square fluctuations (RMSF) are significantly increased in i3 loop upon morphine binding. This increase is visualized in a colored tube diagram (right): larger RMSF values are represented by a thicker backbone and warmer color of the backbone, whereas smaller RMSF values are represented by a thinner backbone and cooler color of the backbone. Although molecular dynamics simulations were performed at the \sim 10 ns time scale, which is short of actual conformational transitions associated with the i3 loop major rearrangements (as previously reported for other GPCRs^{332,333}), the altered dynamics observed by Serohijos et al. points toward critical role of the i3 loop in mediating signal cascade initiated by opiate binding. (Adapted with permission from ref 149. Copyright 2011 Elsevier.)

to changing concentrations of effector metabolites.^{42–46} The alterations in RNA structures upon metabolites binding are so dramatic that they result in complete halt of gene transcription or translation, most likely due to the stabilization of altered riboswitch RNA structure by the metabolite.⁴⁶ For example, in the case of the *Thermotoga petrophila* fluoride riboswitch,⁴⁷ binding of fluoride to this riboswitch induces RNA conformation that is stabilized by the long-range interactions between noncanonical reversed Watson–Crick and Hoogsteen A-U base pairs and the fluoride ion, coordinated by three magnesium ions⁴⁵ (Figure 2B). Significant conformational alterations driven by fluoride are likely responsible for interfering with translation initiation and transcription termination of various genes that encode putative fluoride transporters and enzymes that can be inhibited by fluoride.⁴⁷ Riboswitches represent just a class of RNA molecules that exhibit switch-like behavior upon ligand binding, but the world of RNA is extremely rich and rational RNA regulation will likely emerge as one of the most powerful tools for molecular and cellular biologists.

1.2. Allosteric Regulation

Allosteric regulation is a modulation of protein function in which binding of ligands, post-translational modification, or any other structural excitation leads to conformational changes or concerted motions of structurally distinct protein regions. The binding of an effector molecule at the protein's allosteric site leads to conformational or dynamic changes in the distal active site, affecting its function: effectors that enhance the protein's activity are referred to as *allosteric activators*, whereas those that decrease the protein's activity are called *allosteric inhibitors*. In natural environments, many, if not all, enzymes and signaling proteins are allosterically regulated. For example, ligand-binding at an externally accessible site in G protein coupled receptors (GPCRs) triggers structural changes at distant cytoplasmic domains, which in turn mediate interaction with heterotrimeric G proteins, thereby activating downstream signaling cascades (Figure 3).

Allosteric regulation has been widely employed in various branches of life sciences: from pharmaceutical to basic sciences. Multiple strategies have been devised to control proteins

without directly targeting their active sites.⁴⁸ A major advantage of these strategies, compared to those that perturb the active site directly, is that it offers "noninvasive" control of proteins. Allosteric regulators (i) do not interfere with endogenous regulators that bind to binding sites or its proximity, (ii) can function in concert with direct active site regulators, (iii) often can be more accessible than a buried active site, and (iv) can serve as activatable modules.

In the recent decade, drug discovery has been shifting its focus toward targeting allosteric sites in order to improve compound selectivity.^{49–52} The need for such selectivity is best demonstrated by kinases, which all share a similar ATP binding site, thus making drugs that sterically block ATP binding extremely promiscuous. This promiscuity results in poor specificity and results in drug side effects. Furthermore, allosteric sites are not under evolutionary pressure to be conserved;⁵³ thus, allosteric sites offer unique ligand binding sites with high specificity. Typically, amino acid conservation is associated with its contribution to protein stability, folding kinetics, or a specific function.^{54–56} Amino acids are rather malleable and capable of evolving to bind diverse ligands.⁵⁷ Hence, allosteric sites present a unique and focused opportunity to selectively target one particular protein in a family of homologous proteins.⁵⁸ Allosteric drugs also feature distinct physiochemical properties than their orthosteric counterparts, adding further freedom for discovery of novel active compounds. Furthermore, allosteric drugs can be combined with orthosteric drugs to improve their activities. In summary, allosteric regulation is becoming a new frontier in drug discovery and may become the primary pharmaceutical strategy for many disease targets.

Perhaps one of the most appealing applications of allosteric control is the ability to rationally manipulate specific proteins of interest, as well as their associated cellular pathways. Functional redundancies and nonlinearities in cellular processes underlie inherent complexities of cellular networks, making it challenging to decipher protein functions in the milieu of other proteins and interfering pathways. The ability to control a single protein's function among homologous or functionally duplicating proteins presents an opportunity to glance into cellular wiring at the single protein level. Additionally, such control

should not disrupt the native function of the target protein, thus making allosteric control a unique handle on protein function.

Allosteric control of protein function can also be used for drug target validation of proteins.⁵⁹ Before endeavoring into large scale screening of potential inhibitors of a target protein, it is critical to know whether inhibition is effective. Hence, having an allosteric handle capable of controlling protein function permits direct testing of protein inhibition in endogenous environments, thereby as a proof-of-principle demonstrating the utility of this allosteric site for drug screening.

Understanding the structural and dynamic underpinnings of allosteric regulation and, especially, learning how to engineer allosteric regulation provides an outstanding opportunity to control protein and RNA function and cellular activity. The ability to control cell phenotype through engineered allosteric regulation shows great promise in designing therapies, developing biosensors/biomarkers for early disease diagnosis, and providing a general tool for fundamental biological research.

While allosteric regulation has been a research topic of many laboratories¹⁵ with remarkable progress in understanding how coupling between distant residues is maintained from theoretical (e.g., refs 60–64), experimental (e.g., refs 35, 65–68), and evolutionary (e.g., ref 69) perspectives, we still lack a clear picture of the inter-residue interaction networks that result in allostericity. Here, we review recent advances in our understanding of allostericity and means of controlling protein activities through engineered allostericity.

2. PHYSICAL ORIGIN OF ALLOSTERY

Allostery may seem to be a serendipitous phenomenon that is only inherent to biological molecules. However, the wide reach of allosteric phenomena to virtually all proteins and potentially RNA molecules suggests a possibility that there are deeper origins of this phenomenon. Here, we posit that allostericity is a physical property of many objects that are comprised of a complex network of interactions between constitutive elements. The wiring of these networks allows formation of “channels” within these networks that route the fluctuations from one site of the object to another avoiding dampening of these fluctuations due to numerous other interactions. While not all networks would feature such channels, it is plausible that inter-residue interaction networks that represent molecular structures are necessary for the thermodynamic stability of biological molecules and, for that reason, may be optimized by evolution. Next, we will describe plausible physical and evolutionary origins of allostericity.

2.1. Allostery from a Graph Theory Perspective

Consider a network $G = \{V, E, W\}$ with nodes V representing amino acid (or nucleotide) residues and edges $E = \{\langle e_{ij} \rangle\}$ representing physical interactions between these residues. The weight $W = \{\langle w_{ij} \rangle\}$ of an edge is the strength of the corresponding inter-residue interaction. In the limit when all interactions are homogeneous (equal weight graph, $w_{ij} = \text{constant}$) and local, i.e. each residue is connected to a limited number of nearest neighbors (crystal lattice), any excitation of a particular residue will avalanche through neighbors to other neighbors, thus spreading the signal throughout the whole system and diminishing the impact of this excitation as a function of the distance between the origin and the target site. The same arguments would also hold for homogeneous ($w_{ij} = \text{constant}$) but complete graphs. Hence, the network that would

feature “dedicated” signal transition channels between the origin and the target site is expected to have a mixture of long- and short-range interactions and nonhomogeneous weight distributions in this network.

If we construct graphs based on protein structures, these graphs can be classified as small-world networks,^{70–72} whereas these graphs feature a great degree of clustering like lattices, but much shorter minimal path lengths between the sites than random graphs or lattices.⁷⁰ Here, we survey 5789 proteins taken from the Protein Data Bank (PDB)⁷³ satisfying the following criteria: (i) crystallographic resolution is better than 2.5 Å, (ii) only proteins were surveyed (no DNA, RNA, or other ligands associated with the protein in the crystallographic structure), (iii) no homologous proteins with more than 30% sequence identity were included, and (iv) only the longest chain from every PDB entry was considered (to avoid double counting of monomers in homo-oligomers). For each protein graph we also construct a random graph by random rewiring: we perform random shuffling of the edges between various pairs of nodes, maintaining the same overall amount of edges. We ensure that the distribution of the number of contacts of protein matches that of random graphs. For graph G we define graph neighborhood $R(G)$ of node k to be a graph that contains all nodes that are directly connected to node k , and include all the edges between these nodes that are present in graph G (Figure 4). For each graph we compute its *clustering coefficient*

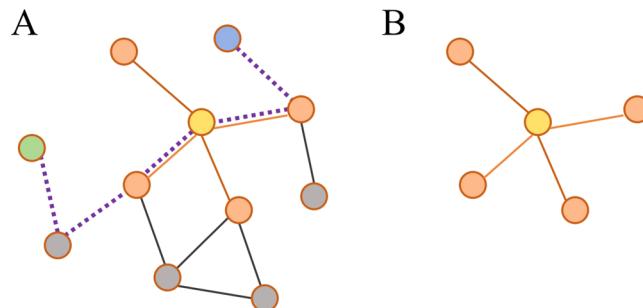


Figure 4. Graphs. (A) Example of a graph with nodes featuring various degrees of connectivity (from one to four edges). The neighborhood of a yellow node is represented by a subgraph with orange edges and nodes (B). The minimal path between the blue and green nodes is depicted by dashed edges.

C , defined as the average fraction of the actual number of connections, n_k , in the neighborhood of a node (k) with N_k neighbors to the total number of possible connections between the nodes in this neighborhood $N_k(N_k - 1)/2$:

$$C = \frac{1}{N} \sum_k \frac{n_k}{\frac{N_k(N_k - 1)}{2}} \quad (1)$$

where N is the number of nodes. We also compute the *average minimal path* L , defined as the average minimal path l_{ij} between any two nodes (Figure 4):

$$L = \frac{1}{N(N - 1)/2} \sum_{i>j} l_{ij} \quad (2)$$

where l_{ij} is the minimal number of edges one needs to transverse in order to travel from node i to node j weighted by the corresponding weights W of the traveled edges.

The clustering coefficients reflect the sizes of neighborhoods, or the reach of a node to other nodes in a given network. From

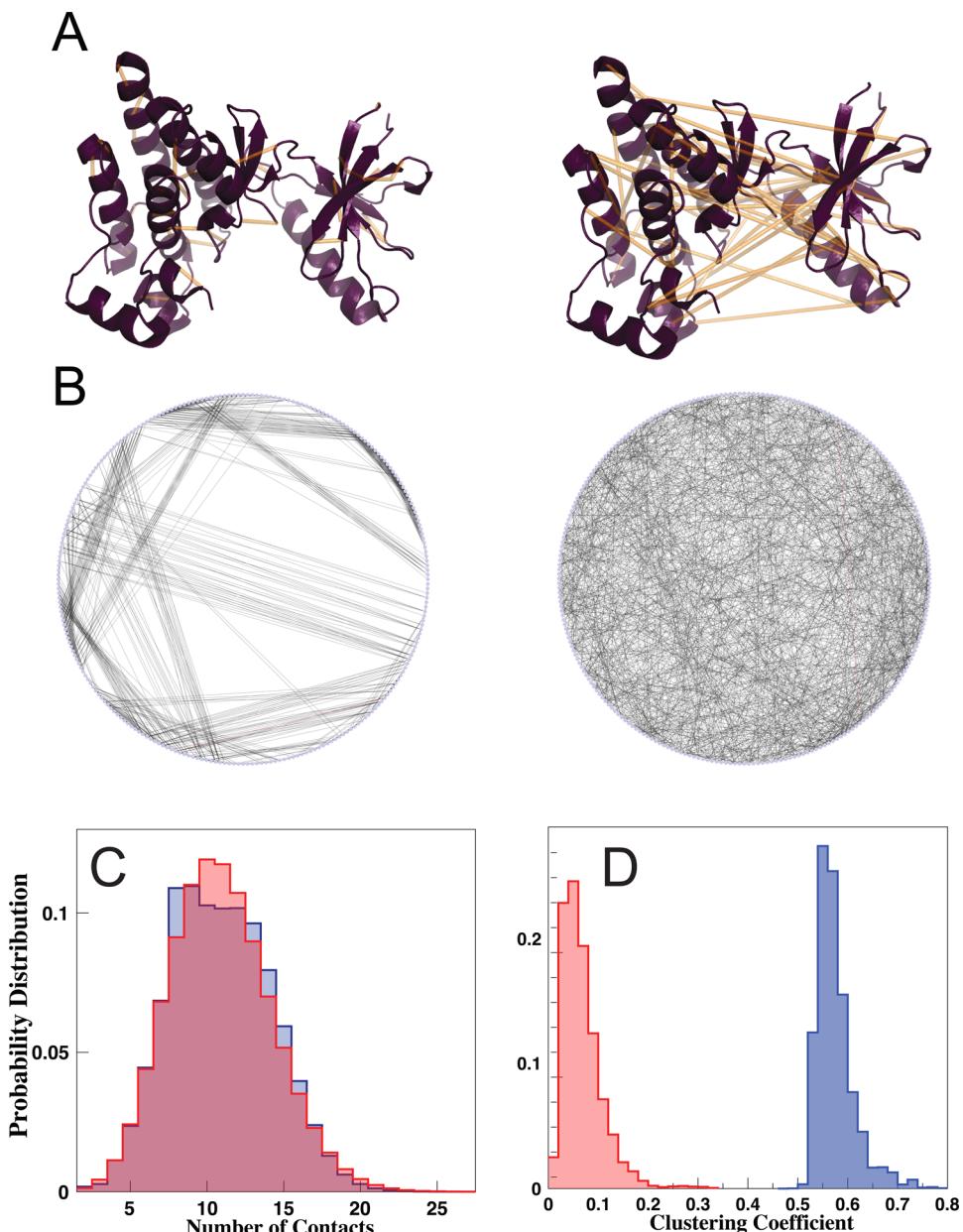


Figure 5. Protein versus random graphs. (A) Overlay of the structure of the focal adhesion kinase (FAK) with the links that connect α carbon atoms located within 8.5 Å (left) versus randomly rewired connections (maintaining the number of connections). For visual representation only a few links are shown on both structures, but the number of links is the same for the FAK graph and the randomly rewired graph. (B) Graph representations of protein interactions (left) versus rewired random graph (right): nodes represent protein amino acids; edges represent connections between any two amino acids whose α carbons are located within 8.5 Å. Despite the dramatic differences between FAK and random graphs, the total number of links in both graphs is the same. (C) Probability distributions of the number of contacts in protein (blue) versus random (red) graphs are indistinguishable from each other (Kolmogorov–Smirnov test $p \approx 0.87$). (D) Probability distributions of the clustering coefficients for protein (blue) versus random (red) graphs are significantly different (Kolmogorov–Smirnov test $p = 10^{-19}$).

in this survey we find that protein graphs feature significantly larger clustering coefficients (Kolmogorov–Smirnov test $p \approx 10^{-19}$) than those of random graphs (Figure 5) and smaller than those of regular lattices.⁷² However, the average minimal paths in protein graphs are much smaller than those of regular lattices. Hence, protein graphs represent a “sweet spot” in the L – C phase diagram, whereby they have the most optimal reach between residues, close to lattices, and have as small an average minimal path as that of random graphs. This “small world” property of protein graphs⁷² makes proteins particularly sensitive to signal reach and propagation across the whole inter-residue communication networks.

Such “small world” protein network organization may originate in the innate hierarchical composition of protein structure: secondary structure allows significant “decrease” of minimal path lengths compared to lattices. For example, β -sheets contribute a large number of long-range through-space interactions. Indeed, for globular proteins it has previously been shown that the probability to form a contact between residues separated by l residues along the protein chain scales as a power-law function of l with the exponent -1.6 .^{54,74}

$$P(l) \sim l^{-1.6} \quad (3)$$

This function is consistent with that for an ideal chain homopolymer, which can be derived from the end-to-end distribution function (taking into account that the number of surface residue interactions is much smaller than that of volume residue interactions):

$$P_H(l) \sim \int_0^{\infty} P(R|l) \theta(R - R_c) dR \sim l^{-3/2} \exp\left[-\frac{3R_c^2}{2l}\right] \quad (4)$$

where R_c is the typical interaction range between amino acids (~ 8 Å between corresponding C_α atoms), and $\theta(x)$ is the Heaviside step function. Both of these functions decay slower than that of any two points on a cubic lattice, which scales as inverse volume enveloped by a sphere of radius l :

$$P_L(l) \sim l^{-3} \quad (5)$$

Hence protein and homopolymer graphs feature identical longer-range connectivity but different than that of lattices.

The heterogeneity of interaction strengths between residues is the primary reasons for the formation of robust allosteric networks in proteins but not in homopolymers. This heterogeneity has a dual impact on the formation of these networks. First, proteins, unlike homopolymers, adapt unique native conformations, thereby presenting a definitive set of pathways within protein graphs that are resilient to structural fluctuations (e.g., due to the finite temperature of the environment), but not in homopolymers. Second, it has been demonstrated that specific inter-residue interaction pathways dictate rapid and specific folding collapse of proteins,⁷¹ resulting in the first-order-like unfolding to folding transition.^{75–79} On the other hand, homopolymers feature a second-order collapse transition that is facilitated by multiple nonunique pathways. The uniqueness of the inter-residue pathways in proteins is governed by heterogeneous interaction strength, which, in turn, is due to evolutionary optimized heterogeneous sequence of amino acids.^{80,81}

While almost any compact object would feature allosteric coupling, like the opposite sides of a Jell-O, proteins are particularly adapted for such coupling due to small-world connectivity within proteins. These allosteric networks may also be strengthened due to evolutionary selection of cooperative, folding-determining inter-residue interaction networks that are robust to external perturbations. Nussinov and colleagues have suggested that all nonfibrous proteins are intrinsically allosteric.²⁸ Protein structural and dynamic properties make them targets for allosteric control of their structural ensembles and function.

2.2. Evolution of Allostery

Globular two-state proteins evolved to rapidly fold into their native functional structure via a nucleation scenario.^{75–77,82–88} Formation of a nucleus is a bottleneck of the folding process and corresponds to the maximum of the free energy along a multidimensional folding reaction coordinate. Such a bottleneck drastically reduces the number of pathways that reach native protein structure, consisting of well-defined contacts formed in the transition state ensemble. Hence, evolution optimized the folding pathways and, in essence, interaction graphs of folding of two-state proteins, and, more generally, those of multistate proteins. Couplings between residues that resulted in folding of globular proteins are quintessential to corresponding protein folds and are present in the native state, as well as in the transition state ensemble. Reichheld et al.

offered support for coupling between evolution of allosteric pathways and protein folding⁸⁹ through the observation that the folding of DNA binding domains and tetracycline binding domains of the homodimeric tetracycline repressor (TetR) are independent in the absence of tetracycline. Upon tetracycline binding to TetR, they observed allosteric coupling of the DNA binding domain which resulted in its folding and binding to DNA. Considering the broader TetR family of proteins, Reichheld et al. argued evolutionary coupling of folding and function, which is potentially transferable to other biological systems.⁸⁹

While the multistate protein folding landscape is more complex than that of a two-state protein, one can apply arguments to multistate folders similar to those mentioned above for two-state folders. We may further argue that coevolution of these interacting residues has resulted in the unique identities of native structures. From this perspective, the notions of structural or dynamic allosteric are inseparable and the difference is rather semantic.

Lockless and Ranganathan recognized that protein allosteric is encoded in protein sequences by evolution, although they have not suggested that allosteric networks may indeed be those responsible for protein folding.⁶⁹ They argued that the thermodynamic impact of covariation of amino acids at two distinct positions could be deduced from the sequences in corresponding positions. Although they used empirical functions to describe this covariation, these functions can be loosely related to physical functions describing thermodynamics of the protein sequence space and, thus, have physical meaning. Lockless and Ranganathan were able to successfully predict and experimentally verify allosterically coupled sites in the PDZ domain family.

Ranganathan and colleagues have further used this approach to identify allosteric coupling of residues in members of the G protein superfamily, which contain GTP/GDP dependent switches responsible for interaction specificity with downstream binding partners.⁹⁰ Using their statistical coupling approach, they detected allosteric coupling between residues responsible for nucleotide binding and protein–protein interactions, suggesting that an “allosteric core” underlies different G protein family members. Importantly, from this and other works Ranganathan and colleagues demonstrated that the allosteric coupling underlies both folding and function of proteins.^{90–92} Similar suggestions were also made by Nussinov et al.^{93,94} and, using HSP70 as an example, Zhuravleva et al.⁹⁵

2.3. Identifying Allosteric Pathways

The end of the last century and the beginning of this one brought many revolutionary breakthroughs in our understanding of how proteins fold into their native structures and how their structures evolve. These breakthroughs enabled scientists to apply this knowledge to actually control and manipulate different aspects of protein lives. Some of the most outstanding advances came from the ability to manipulate protein-folding kinetics^{96,97} and design new proteins.^{98–101} The ability to design and manipulate protein allosteric has also recently enjoyed significant advances, which will be described in the next sections. Here, we discuss means to determine protein allosteric networks.

2.3.1. Physics-Based Approach. Unlike in homopolymers, protein structure is closely linked to folding dynamics. This relationship is ultimately responsible for self-organized¹⁰² allosteric networks of long-range inter-residue interactions. A

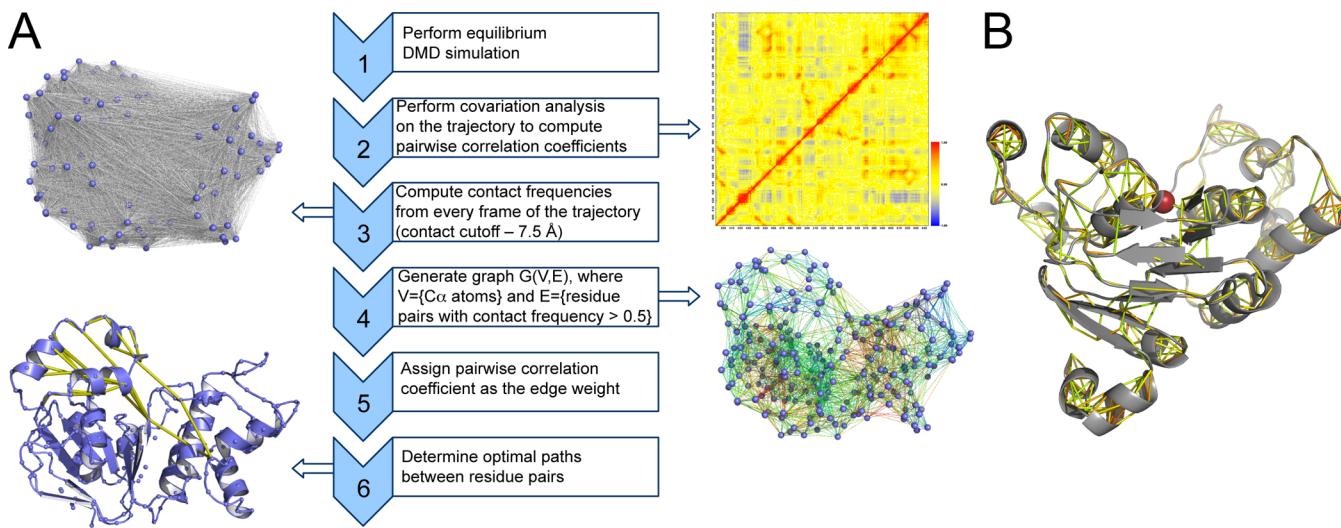


Figure 6. Mapping allosteric networks in proteins. Allosteric networks provide guidance in selection of critical residues. (A) Flowchart outlines steps involved in determination of optimal paths between chosen sites in a given protein. (Reprinted with permission from ref 116. Copyright 2015 The Royal Society of Chemistry.) (B) Pruned correlated motion network in NBD1 (cutoff = 0.68) overlaid on structural representation. Strength of correlations ranges from weak (green) to strong (red), with critical residue S492 represented as a red sphere.

graph-theoretical or network approach^{70,103,104} thus seems a natural choice for identifying allosteric networks.^{48,105,106} Indeed, with the network approach there has been substantial effort to describe protein folding pathways,^{107,108} critical residues (nuclei)^{77,84,85,109–111} responsible for crossing of the transition state barrier en route to its native state,^{71,72,112,113} as well as allosteric pathways linking specific sites on a protein for binding of potential effectors.^{114–116}

The focal concept underlying the network approach in applications targeting identification of allosteric pathways is that effector binding, light activation, or posttranslational modifications result in a propagating cascade of residue fluctuations throughout the structure to the active site of the protein. By measuring these fluctuations, e.g. using molecular dynamics simulations or normal-mode analysis, these fluctuations or correlated motions are mapped onto a graph with nodes representing residues (or groups of residues) and edges representing weights of measured dynamic properties of the system.

Molecular dynamics simulations provide invaluable information about systems dynamics at atomic resolution. Although the force field that is used to guide simulations is an approximation to interatomic forces and, hence, not fully accurate, recent success in simulations of folding processes provides validation to molecular dynamics simulation approaches. From molecular dynamics simulations, Chen et al.,¹¹³ Chen et al.,¹¹⁷ Sharma et al.,¹¹⁸ and Teotico et al.¹¹⁹ used a covariance matrix¹²⁰ to determine a coupling matrix between protein residue pairs:

$$\text{cov}(i, j) = \langle \vec{e}_i - \langle \vec{e}_i \rangle \rangle \langle \vec{e}_j - \langle \vec{e}_j \rangle \rangle \quad (6)$$

where \vec{e}_i is the unit vector of the displacement of the α carbon j . This approach allows one to determine direct coupling between any pair of residues.

In order to identify the path that connects these pairs of residues, Proctor et al.¹¹⁶ proposed to treat the normalized covariance matrix as a weighted graph connectivity matrix, with the weights of edges $E_{ij} = 1 - |\text{cov}(i,j)|$ (Figure 6). This graph is likely a complete but heterogeneously weighted graph. In order to extract information about dominant (optimal) paths,^{121,122}

Proctor et al.¹¹⁶ first pruned the network by eliminating nonphysical interactions, i.e. between residue pairs that do not maintain physical contacts, or the distance between β carbons of corresponding atoms (α carbons for glycines) is greater than 7.5 Å, is maintained less than 50% of simulation time. Then, the authors imposed a cutoff E_c , such that any edge is disconnected if its value less than E_c . The value of the cutoff E_c is chosen in such a way that 50% of all edges belong to the disconnected subgraphs and the graph becomes critical.^{80,121} This approach leaves no free parameters in the system. Using this procedure, the authors emphasized the propagation of local correlations in motions throughout the network. Proctor et al. then used the Dijkstra algorithm¹²³ to identify the optimal path responsible for propagation of correlations through the network. An application of the optimal path-mapping algorithm is described in section 5.1.

Besides identifying pathways within protein graphs, Proctor et al. suggested an approach to determine the criticality of a specific residue (node) to connectivity of the largest component.⁹⁵ Critical residues or “bottlenecks” disconnect the largest subgraph, thereby paralyzing communications between the active and allosteric sites. The importance of bottleneck residues is difficult to underestimate: its small perturbation may result in a catastrophic impact on the network topology. Hence, these residues can also serve as targets for modulation. Applications of the network approaches have been successfully utilized in studies of cystic fibrosis (CF) associated protein cystic fibrosis transmembrane conductance regulator (CFTR) (section 5.1). Sethi et al. proposed to partition protein graphs, constructed based on the dynamic inter-residue coupling, into subgraphs they termed “communities”.¹²⁴ By analyzing coupling between communities of interacting residues in archaeal leucyl-tRNA synthetase (LeuRS):tRNA^{Leu} and bacterial glutamyl-tRNA synthetase (GluRS):tRNA^{Glu} complexes, Sethi et al. found that both complexes featured remarkably similar organizations of the communities, and that these communities were loosely coupled to each other, while communities themselves represent residues that are strongly coupled to each other. This finding is consistent with the small-world properties of protein structural networks.^{71,72}

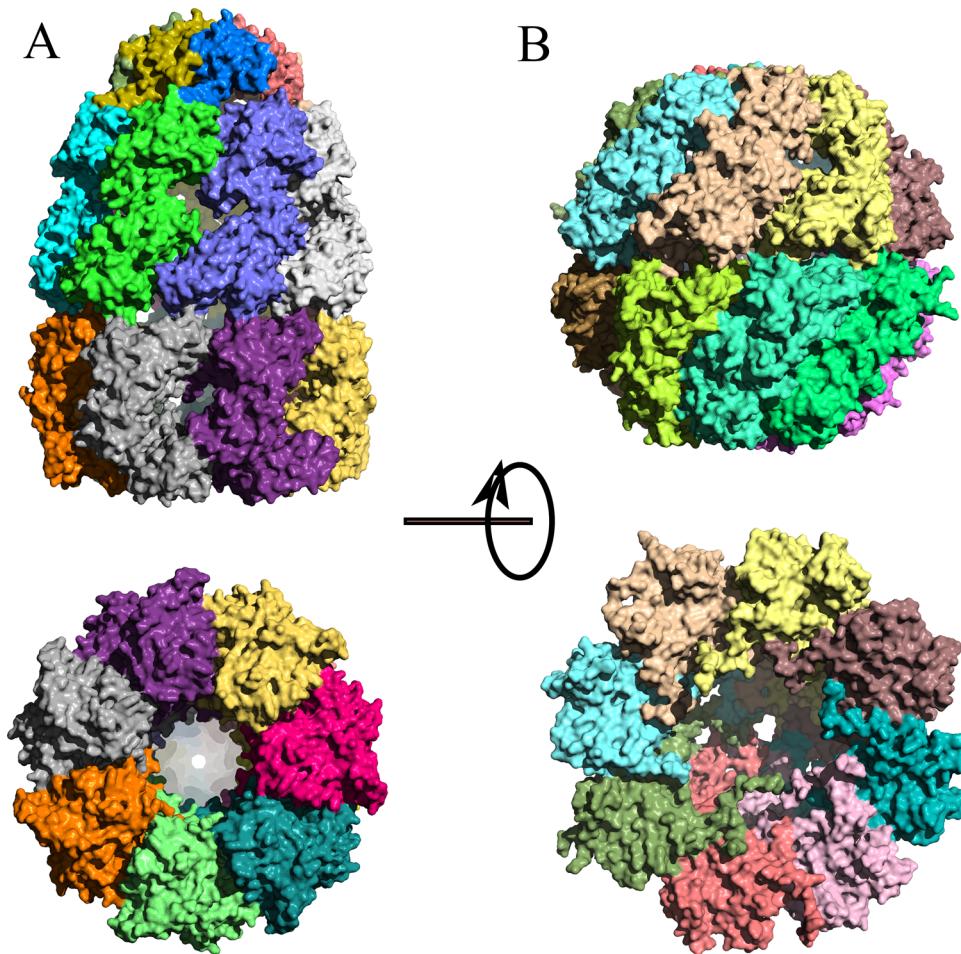


Figure 7. Chaperonins. (A) Structure of the group I chaperonin from *E. coli*, GroEL–GroES complex.¹⁷⁵ (B) Structure of the group II chaperonin CCT¹⁸¹ from yeast.

Graph theory has proven to be a well-suited framework for understanding connections between remote entities. The above-described approach presents a powerful yet relatively computationally inexpensive approach to uncover allosteric connectivities of residues in proteins. One can extend these approaches to intermediate states of proteins, which, in turn, can be mapped using graph theoretical approaches.¹⁰⁸ For example, Serohijos et al. proposed to map folding pathways and intermediates of protein folding by building a protein state graph, in which a node represents a state-representative protein conformation, the edges between two nodes represent physical transitions between states, and the weights of the edges are actual probabilities of transition between states.¹⁰⁸ Using state graphs, Serohijos et al. identified distinct states and pathways responsible for CFTR's nucleotide binding domain 1 (NBD1) misfolding in mutant but not wild type protein.¹⁰⁸ Identification of critical intermediates allows one to glance at the states that control protein function and use allosteric approaches to control these states.

An alternative to covariation analysis was proposed by Bhattacharya and Vaidehi,¹²⁵ who combined accelerated molecular dynamics simulations and pairwise time-averaged mutual information between residues (MI) computed in internal coordinates. They constructed a graph based on the MI matrix as the weight matrix of the complete graph, and then pruned the edges with a specifically chosen cutoff value. Based on the constructed graph, they studied propagation of the

allosteric signal within the GPCR β_2 -adrenergic receptor (β_2 AR) in response to the inverse agonist and the agonist in various conformational states of β_2 AR. By clustering optimal pathways, the authors were able to detect “allosteric pipelines” connecting extracellular and intracellular portions of this protein, thereby offering an insight into the transmembrane communication of β_2 AR leading to G-protein activation and downstream signaling.

One of the most critical challenges of the network approach is that any dynamic signatures obtained from the system using computational or experimental tools reflect motions that occur at time scales of actual conformational change associated with allosteric coupling. Hence, assessment of sampled time scales by either computation or experiment is necessary for delineation of allosteric pathways using a network approach. A remarkable application of this approach was described by Gasper et al.¹¹⁴ in studies of thrombin. In their studies, the authors employed accelerated molecular dynamics (AMD), pioneered by Andrew McCammon,^{126,127} to access time scales of protein motion that contribute to large scale conformational changes. Amaro et al. used steered molecular dynamics (SMD)^{128–131} to delineate allosteric pathways in glutamine amidotransferase.¹³²

Proctor et al. described a network approach and applied it to cystic fibrosis transmembrane conductance regulator (CFTR), mutations of which cause cystic fibrosis, using discrete molecular dynamics (DMD) simulations^{133–139} (section

5.1).¹¹⁶ The DMD algorithm enables the extensive sampling required for simulations of large conformational changes,¹⁴⁰ protein folding,^{141,142} and protein–peptide interactions^{143,146} at a reasonable computational time scale. Like AMD, the DMD algorithm significantly extends the time scales reachable by molecular dynamics, enabling elucidation of the functions of such large molecules as CFTR,^{116,147} dynein,¹⁴⁸ μ -opioid receptor,¹⁴⁹ and epithelial sodium channel.¹⁵⁰

Another approach to uncover coupled motions at the time and length scales required for understanding protein function is by using normal-mode analysis, which is used to examine protein vibrations around a minimal energy structure by a quadratic approximation of the potential energy.^{151–157} Due to quadratic approximation, the method of normal-mode analysis requires diagonalization of the Hessian matrix with dimensions $3N \times 3N$, where N is the number of particles to consider. For large proteins, the diagonalization of the Hessian matrix may become limiting due to higher matrix rank. Several approaches have been proposed to circumvent this limitation.^{158–161} Normal-mode analysis has also been simplified using the elastic network model.^{162,163} In this method particles within a certain cutoff distance are considered to be connected by springs, and modeled by a simple Hookean potential. Bahar and colleagues further developed the Gaussian network model.^{164,165} In this model, residues within a given cutoff distance are connected by a spring with a uniform spring constant. The Gaussian network model provides information on residue mean square fluctuation and cross-correlations between residues in individual modes. However, it cannot provide information on three-dimensional motions as it uses an $N \times N$ Kirchoff matrix rather than the $3N \times 3N$ Hessian, where N is the number of residues. The anisotropic network model^{166,167} uses the Cartesian coordinates of C_{α} atoms to construct the Hessian matrix, and is essentially a coarse-grained normal-mode analysis similar to that originally constructed by Hinsen.¹⁶³ Normal-mode analysis has been widely employed for detecting allosteric sites^{168,169} and mapping allosteric pathways^{124,170,171} in biological molecules.

Normal-mode analysis had a revolutionary impact on the structural biology field as it allowed studies of large-scale cellular machinery. Chaperonins^{172–174} are critical elements of cellular machinery that, predominantly by utilizing ATP hydrolysis, assist protein folding by disrupting non-native amino acid interactions and allow newly synthesized proteins to form their native structures. There are two classes of chaperonins, group I and group II: Group I chaperonins are typically present in mitochondria, eubacteria, and chloroplasts, while group II chaperonins are typical to archaea and eukaryotes. While the difference between these classes is unambiguous—distinct structural organization and stoichiometry of subunits—a bird's eye view of these proteins reveals barrel-like structures (two double rings). The barrels feature large hydrophobic patches exposed to the solvent, thereby allowing to chaperonins to interact with misfolded proteins, and also exposing hydrophobic surfaces.

Surprisingly, despite strong similarities between domains comprising group I and group II chaperonins, these two classes represent distinct allosteric properties. For example, allosteric pathways of GroEL,¹⁷⁵ the group I chaperonin from bacteria (Figure 7A), have been thoroughly mapped using a battery of biochemical approaches,^{176,177} and further corroborated by computational studies using normal-mode analysis by Ma and Karplus.¹⁷⁸ These studies revealed that GroEL undergoes

concerted intraring transitions that is well-fit by the Monod–Wyman–Changeux model.²⁰

Unlike GroEL, thermosome and TRiC^{179–181} (TCP-1 ring complex, also known as chaperonin containing TCP-1 (CCT)) (Figure 7B) follow sequential allosteric interdomain transition,¹⁸² consistent with the Koshland–Nemethy–Filmer model.¹⁸³ Additional computational studies^{184–188} based on normal-mode analysis have further mapped the intraring allosteric networks and communication across subunits, further supporting the allosteric mechanisms proposed by the Horovitz laboratory for these two classes of chaperonins. These studies have been extremely important not only to our understanding of how protein folding machinery works in cells, but also offered plausible pharmaceutical strategies to target protein misfolding. These studies also demonstrated the power of normal-mode analysis to significantly extend the time scales currently available to molecular dynamics simulations, and glance into actions of large molecular complexes.

2.3.2. Evolutionary Approach. Lockless and Ranganathan proposed an approach, statistical coupling analysis (SCA), to detect allosteric coupling in proteins using information about amino acid covariation in protein families using multiple sequence alignment (MSA).⁶⁹ The authors proposed an empirical evolutionary conservation parameter at a position i along the protein sequence in MSA:

$$\Delta G_i^{\text{stat}} = kT^* \sqrt{\sum_x \left(\ln \frac{P_i^x}{P_{\text{MSA}}^x} \right)^2} \quad (7)$$

where kT^* is an arbitrary energy unit, P_i^x is the probability of observing amino acid x from the multiple sequence alignment, and P_{MSA}^x is the frequency of observing amino acid x anywhere in the sequences. The authors further defined empirical coupling energy $\Delta\Delta G_{ij}^{\text{stat}}$ between sites i and j as

$$\Delta\Delta G_{ij}^{\text{stat}} = kT^* \sqrt{\sum_x \left(\ln \frac{P_{il\delta j}^x}{P_{\text{MSA}l\delta j}^x} - \ln \frac{P_i^x}{P_{\text{MSA}}^x} \right)^2} \quad (8)$$

where $P_{il\delta j}^x$ and $P_{\text{MSA}l\delta j}^x$ represent corresponding probabilities but in a subset of MSA that represents a perturbation at site j . Using this measure, Lockless and Ranganathan detected coupling of distant residues in PDZ domains (postsynaptic density of neuronal excitatory synapses (PSD-95), discs large, zona occludens 1).¹⁸⁹ Mutagenesis confirmed unexpected long-range interaction between residues identified through evolutionary analysis. The SCA approach has been used by the Ranganathan group to uncover long-range allosteric communication in G protein coupled receptors (GPCR),¹⁹⁰ guanine nucleotide proteins G_{Saw} ,⁹⁰ S1A serine proteases,⁹¹ dihydrofolate reductase,⁹² and the cysteine peptidase cathepsin K.⁶²

The SCA approach was further extended and generalized by the Thirumalai group.^{113,191} The authors made less restraining assumptions about the distributions of amino acids at each position along the protein sequence in MSA. They used augmented SCA to test communication networks in the PDZ domain family, GPCRs, lectins, and dihydrofolate reductase.

The SCA approach is one among several established algorithms that are based on evolutionary information derived from the MSA:¹⁹² observed minus expected squared (OMES) covariance algorithm,¹⁹³ mutual information (MI) covariance algorithm,¹⁹⁴ and McLachlan based substitution correlation (McBASC).^{195,196} Interestingly, these approaches do not fully

agree with each other, suggesting sensitivities to different aspects of amino acid conservation during evolution.¹⁹² Fodor and Aldrich have argued that the thermodynamic coupling is not limited to evolutionary constraint residues,¹⁹⁷ yet the information about residue covariation in MSA proved to be a powerful predictor of protein structure¹⁹⁸ for some proteins, as multiple groups independently developed methods for protein structure determination using covariation analysis.^{199–201}

Principal limitations of SCA and other approaches lies in the strong signal dependence on the number of available family members in MSA, noise due to common ancestry,^{202,203} and transitivity.^{198,199} The problem of limited number of sequences in MSA might potentially be overcome by “designing” protein family members using computational protein design algorithms.²⁰⁴

Perhaps the most critical assumption pertaining to SCA methodology is that nature imposes evolutionary pressure to preserve allosteric pathways in proteins. This assumption has been challenged by many independent studies, bringing into controversy the success attributed to SCA methodology. Previous work by Hudson et al.²⁰⁵ demonstrated that allosteric networks in proteins are not conserved in the course of evolution, and, in the case of phosphorylases, three distinct groups of networks exist within these proteins. A prominent example of divergent allosteric pathways is the family of GPCRs: depending on the effector molecules, which bind to the same active site with minor variations, the signal propagating to the G-protein complex varies significantly resulting in radically different phenotypes.^{149,206–209} Livesay et al.²¹⁰ offered the example of hemoglobin, whose allostery has been under scrutiny for decades. Despite the high structural similarity between hemoglobin family members, a plethora of distinct allosteric pathways exist that relate to their function.^{211,212} Furthermore, according to Kuriyan and Eisenberg, the malleability of these allosteric pathways is a mechanism for functional diversification during the course of evolution.²¹³ Hence, the success of the SCA approach is surprising in light of the unsubstantiated fundamental assumption.

In pathway-mapping studies performed in my laboratory, we have often found multiple allosteric pathways within proteins.¹¹⁶ In agreement with our experience, del Sol et al.,²¹⁴ Cui and Karplus,^{40,215} and Sethi et al.¹²⁴ have all confirmed the multiplicity of allosteric pathways in proteins. Hence, the success of SCA is potentially rather incidental.

From bioinformatics studies, Fodor and Aldrich noted that correlated mutations are mostly detected between residues that are in physical proximity.¹⁹⁷ Liu et al. corroborated the findings of Fodor and Aldrich using theoretical studies, in which they used an exactly solvable two-dimensional lattice model of a protein to explicitly compute residue coupling in this model and relate it to SCA.²¹⁶ They found that only physically proximal residues are identified by SCA, further supporting the earlier findings by Fodor and Aldrich.¹⁹⁷ Hence, one possibility for the success of SCA is that considered proteins are small globular proteins with a well-defined hydrophobic core, which tend to be conserved in the course of evolution.⁵⁴ Due to the multiplicity of allosteric pathways,¹²⁴ conserved residues that constitute the hydrophobic core of a small protein are likely to reside along allosteric paths. Indeed, Tang et al. have demonstrated that one can predict allosteric pathways in myosin by simply following a path of conserved residues.²¹⁷

More validation is required, including negative controls, for the SCA approach to be broadly applicable to diverse proteins.

2.3.3. Nuclear Magnetic Resonance. Among multiple approaches to measuring allostery in proteins, nuclear magnetic resonance (NMR) stands out due to its ability to probe protein dynamics in a broad range of time scales: from picosecond to nanosecond dynamics of side chains and from microsecond to millisecond dynamics of the backbone.^{36,37,218–225} In order to probe allosteric coupling at multiple scales, Clarkson et al.²²⁶ studied the small protein eglin c, which has previously been considered to be a nonallosteric protein. By combining mutagenesis (valine to alanine) with NMR studies, such as spin relaxation, residual dipolar couplings, and scalar couplings, the authors were able to probe the response of the mutations on the protein dynamics and map the allosteric network within eglin c. They found that perturbation due to mutagenesis spreads throughout eglin c altering both the dynamics and structure of this protein, spanning distances of ~16 Å,²²⁶ thereby negating a previous belief that eglin c is a nonallosteric protein. Using similar approaches, Fuentes et al. interrogated allosteric coupling within PDZ domains and measured the impact of critical residues on the coupling network.²²⁷

One of the most direct determination of allosteric pathways was offered by Selvaratnam et al.⁶⁶ The authors performed a covariance analysis of NMR chemical shifts, clustering, and singular value decomposition (termed CHESCA) to reveal details of allosteric coupling throughout the protein EPAC, cAMP-binding domain of the protein, which is activated by cAMP. They found that an extensive allosteric network throughout EPAC that spans distant regions, including the effector binding site, and confirmed their studies through site-directed mutagenesis. In particular, studies by the Lee and Melacini groups exemplify some of the most direct approaches to detecting protein allostery using biophysical and biochemical techniques and NMR.

Direct protein structural determination using NMR is time-consuming and laborious. Combining less laborious NMR approaches with molecular dynamics simulations offers opportunities to observe allosteric changes in biological molecules upon effector binding. While applications of these approaches to allosteric pathway mapping have yet to be fully established, they present outstanding opportunities for such purposes. The Vendruscolo laboratory proposed an elegant method for incorporating chemical shifts,^{228–231} residual dipolar coupling,²³² and the hydrogen exchange residue protection factor²³³ as parabolic restraining functions that sum to the force field E_0 :

$$E = E_0 + w \sum_{i=1}^N (x_{\text{exp}} - x_{\text{sim}})^2 \quad (9)$$

where x_{exp} and x_{sim} are experimental and instantaneous simulation values of a particular observable, measured in a given experiment; w is the weight of the added term; and N is the number of experimentally determined constraints. This approach has been proven to determine protein conformational ensembles for test systems. One limitation of this approach is that it strongly depends on the number of experimental constraints N . In the underdefined system ($N \ll L^\alpha$, where L is the number of amino acids in the given protein and $\alpha = \{1, 2\}$ is a function of the method used) it may be impossible to delineate specific structural ensembles due to the smaller and noisier second term of eq 9.

An alternative approach to utilizing hydrogen exchange residue protection NMR experiments was offered by Dixon et al.,²³⁴ who proposed a methodology to decompose protection factors into a pairwise matrix of inter-residue interactions, which then were added in a similar fashion to eq 9 to the “naïve” force field E_0 . Such an approach allowed mapping of rare yet functional conformations of the focal adhesion kinase targeting domain, which is responsible for the formation of focal adhesions in living cells.

Chen et al. suggested omitting any addition to the “naïve” force field during simulations, and instead to perform unbiased simulations of the system.²³⁵ Chen et al. assumed that, during the course of sufficiently long simulations, we recover representative conformations that satisfy experimental constraints. To recover these conformations, the authors minimized the objective function:

$$\chi^2 = \frac{1}{N} \sum_{i=1}^N (S_{i,\text{exp}}^2 - S_{i,\text{sim}}^2)^2 \quad (10)$$

where $S_{i,\text{exp}}^2$ are experimentally determined order parameters, while $S_{i,\text{sim}}^2$ are calculated order parameters from unbiased simulations. While this approach, termed *sample and select*, strongly depends on the sampling efficiency of molecular dynamics simulations, it offers an unbiased view of molecular ensembles consistent with the given order parameters.

2.3.4. ϕ -Value Analysis. An intriguing approach to map allosteric pathway was proposed by Yifrach and Horovitz¹⁷⁷ and later by Auerbach and co-workers.^{236,237} Although their approach is limited due to the number of stringent assumptions, it offers a straightforward method of determining allosteric pathways for some proteins. The most important yet limiting assumption is that two allosterically linked states exist, these states are separated by a free energy barrier (Figure 8), and a single path exists to cross that barrier. With these assumptions, they utilize Leffler’s linear ansatz,²³⁸ which relates the changes in free energy of the transition state ΔF^\neq to that of the two states ΔF_1 and ΔF_2 :

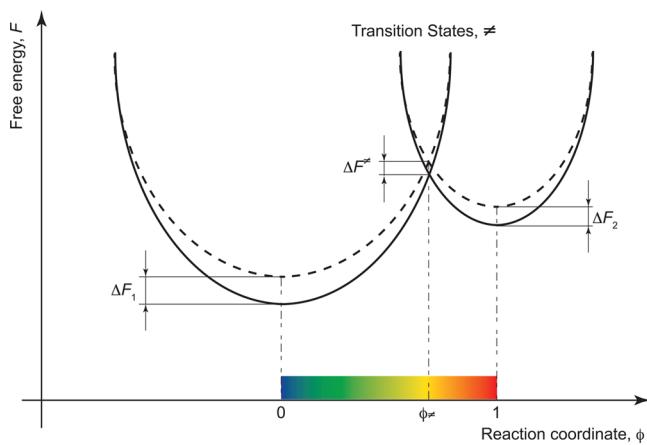


Figure 8. ϕ -value analysis. In a simplified allosteric model, two allosteric states are separated by a free energy barrier (transition state ensemble, \neq). Leffler’s ansatz parameter ϕ ($0 \leq \phi \leq 1$) is a reaction coordinate that serves as a guide between the two states. Through perturbation of a protein, it is possible to determine the impact of various residues on the transition state ensemble, and thereby map the protein’s allosteric pathways.

$$\Delta F^\neq = \phi \Delta F_1 + (1 - \phi) \Delta F_2 \quad (11)$$

where $0 \leq \phi \leq 1$ is the reaction coordinate parameter that controls the separation of the two states from each other and from the transition state. Equation 10 can be rewritten in terms of the kinetic rates of passing the barrier and the equilibrium rate constant between the two states, K_{eq} :

$$\ln k^\neq = \phi \ln K_{\text{eq}} + C \quad (12)$$

where C is some constant. By perturbation of K_{eq} via mutagenesis followed by measuring changes in the reaction rates (passing through the transition state barrier), we can estimate the impact of the residue to the transition state:

$$\phi = \partial \ln k^\neq / \partial \ln K_{\text{eq}} \quad (13)$$

Therefore, by performing extensive mutagenesis for each site along the polypeptide chain, we can identify those residues with the largest contribution to the transition state pathway and map the allosteric path between the two states.

This approach has been extensively utilized in the protein folding field,^{84,96,110,239–243} also known as ϕ -value analysis. Using ϕ -value analysis, the transition state ensemble of a two-state protein is determined by probing the impact of each residue via mutagenesis. Originally pioneered by Alan Fersht, this approach became a gold standard for probing the transition state of two-state protein folding pathways.^{75,244,245} As in two-state protein folding, if protein allostery is governed by a single barrier, ϕ -value analysis offers a direct approach to delineate this path.

Grosman et al.²³⁶ applied ϕ -value analysis to heteromeric acetylcholine receptor, an ion channel that controls synaptic transmission by modulating open, or ion-permeable, and closed, or ion-impermeable, states. To determine the rates of opening and closing and the equilibrium constants between the open and closed states, the authors performed single-channel current measurements for a series of mutants. They found a well-defined single path of residues responsible for switching between the open and closed states of acetylcholine receptor. This approach has also been applied to other allosteric systems, such as ATP-dependent allosteric conformational transition of the GroEL chaperonin^{177,246} and functioning of a Cl^- ion channel CFTR.²⁴⁷

The applicability of ϕ -value analysis is critically compromised by the assumption of a single path connecting the two states (Figure 9). In reality, the majority of proteins feature multiple allosteric pathways,^{40,214,215} rendering this method of analysis limitedly applicable. A further critical limitation of ϕ -value analysis is the assumption of structural invariance upon residue perturbation (eq 13). Delineation of the scale of structural alterations following amino acid substitutions and their impact on the transition states is often infeasible. In the protein folding field, the application of ϕ -value analysis for two-state proteins has been additionally limited due to the complexity and cost of the required experiments.²⁴⁰

2.3.5. Double Mutant Cycles. Fuentes et al.,²²⁷ Halabi et al.,⁹¹ and Hultqvist et al.²⁴⁸ proposed to measure residue coupling through double mutant cycles,²⁴⁹ whereby two probed residues X and Y are mutated to alanine. Fuentes et al.²²⁷ measured corresponding changes in the unfolding free energies of mutant protein $\Delta G(\{X,Y\} \rightarrow A)$ with respect to wild type $\Delta G(\text{WT})$:

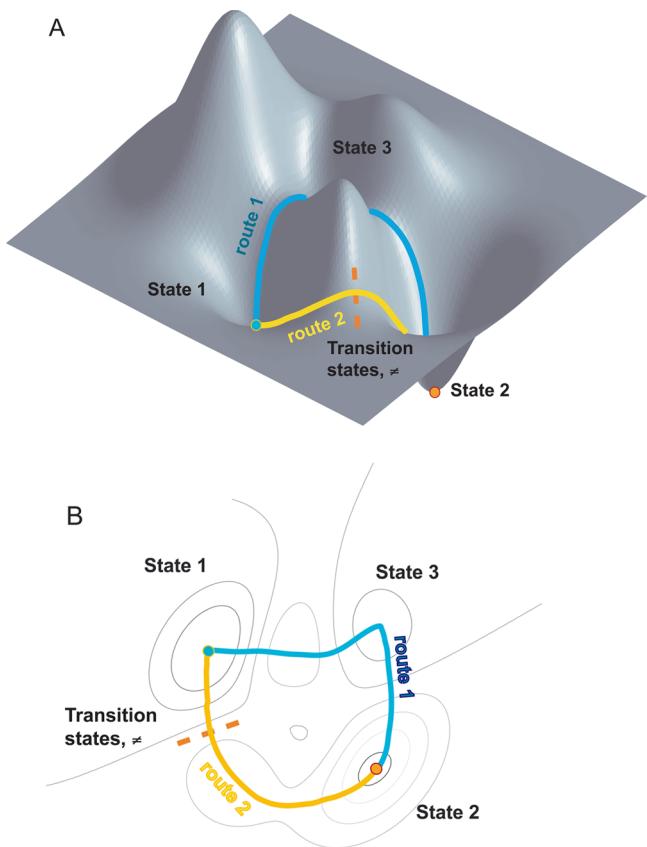


Figure 9. Multiplicity of allosteric pathways. (A) Three-dimensional and (B) contour plots of the free energy landscape of a hypothetical protein featuring two allosteric pathways connecting states 1 and 2, either through another state 3 (more likely path due to lower barriers) or a direct path (less likely path). (Adapted with permission from ref 79. Copyright 2006 Elsevier.)

$$\Delta\Delta G(\{X, Y\} \rightarrow A) = \Delta G(\{X, Y\} \rightarrow A) - \Delta G(WT) \quad (14)$$

Fuentes et al.²²⁷ then introduced the double mutation $X \rightarrow A$ and $Y \rightarrow A$, where the net change in free energy of the double mutant with respect to the corresponding changes due to each individual substitution is

$$\begin{aligned} \Delta\Delta\Delta G(X \rightarrow A, Y \rightarrow A) &= \Delta\Delta G(X \rightarrow A, Y \rightarrow A) \\ &\quad - (\Delta\Delta G(X \rightarrow A) + \Delta\Delta G(Y \rightarrow A)) \end{aligned} \quad (15)$$

If residues are independent and uncoupled, then we expect that $\Delta\Delta\Delta G(X \rightarrow A, Y \rightarrow A) = 0$ as there is no interaction between residues X and Y. Hence, $\Delta\Delta\Delta G(X \rightarrow A, Y \rightarrow A)$ measures the degree of coupling between residues X and Y.

Similarly to Fuentes et al.,²²⁷ Halabi et al.⁹¹ measured the change in the melting temperature ΔT_F or catalytic activity $\Delta(k_{cat}/K_M)$ of double mutants, following a formalism similar to that described in eqs 14 and 15.

While the approach proposed by Fuentes et al.,²²⁷ Halabi et al.,⁹¹ and Hultqvist et al.²⁴⁸ offers a direct measurement of the coupling between residues, it rests on the fundamental assumption that the structural ensembles of mutant and double mutant proteins remain similar to those of the wild type protein. This assumption is unlikely to strictly hold for even very stable proteins. Furthermore, the typical values of $\Delta\Delta\Delta G(X \rightarrow A, Y \rightarrow A)$ are close to zero and feature compounded error bars from the several required ΔG

measurements. Hence, the accuracy of these coupling free energies and whether they reflect actual inter-residue coupling in proteins remains unclear.

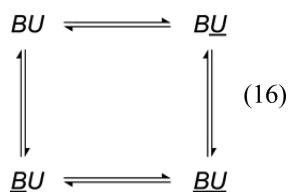
2.4. Identifying Allosteric Sites Using Small Molecules Screening and Crystallography

Perhaps the most direct approach to identify allosteric sites is via small molecule screening and crystallographic resolution of complex structures. When contrasting apo and holo structures of proteins, one can select specific compounds that bind to proteins and allosterically alter the active site. The advantage of this approach is that it allows identification of the allosteric compounds and visualization of conformational changes or changes in the dynamics of the active site (by contrasting B-factors). Furthermore, several groups have utilized this approach and discovered a substantial number of allosteric drugs targeting HIV reverse transcriptase,²⁵⁰ glycogen phosphorylase,^{251,252} p38 kinase, and others.²⁵³ An ingenious variation of this approach was offered by Hardy et al.,²⁵⁴ who screened a library of 10 000 thiol-containing compounds against caspases 3 and 7. In their screen, they identified a cysteine residue located in a deep pocket that, upon disulfide cross-linking with a small molecule ligand, triggered substantial conformational change of the active site of the proteins and inhibited their function. Such approaches are very powerful, but very resource-intensive and expensive, making them mostly accessible in industrial(-like) settings.

3. CONTROL OF ALLOSTERY USING LIGANDS

3.1. Protein Activity Regulation through Mutually Exclusive Folding

One of the original regulatable protein constructs was proposed by Radley et al.,²⁵⁵ who constructed a bifunctional fusion protein that consisted of two unrelated, independently folding domains. Due to steric hindrances at the junction site, these proteins cannot fold simultaneously, causing one protein to be preferentially folded, while leaving the other one unfolded. Upon temperature change, the thermodynamic equilibrium between states shifts, resulting in alternate populations of folded domains within the construct. Hence, by regulating the temperature, Radley et al.²⁵⁵ were able to construct a regulatable protein. As proof-of-principle, the authors chose to insert ubiquitin (U) into a surface loop of ribonuclease barnase (B) from *Bacillus amyloliquefaciens* (Figure 10). The distance between N- and C-termini of ubiquitin (38.5 Å) greatly exceeds the distance between N- and C-termini side loop ends (10.4 Å) of the cut loop from the barnase. Simultaneously forcing ubiquitin termini into the loop of barnase is sterically incompatible with both folded structures, making barnase more dominant at lower temperatures within the synthesized construct (UB). Let us denote by U, \underline{U} , B, and \underline{B} folded and unfolded states of ubiquitin and barnase domains, respectively. Then the UB construct coexists in four states, UB, $\underline{U}B$, $U\underline{B}$ and $\underline{U}\underline{B}$. At lower temperatures ($T < 29$ °C), $\underline{U}B$ state is the most dominant, at $T = 29$ °C both $\underline{U}B$ and $U\underline{B}$ states are equally populated, and at higher temperature (29 °C $< T < 90$ °C), $U\underline{B}$ state is the most dominant. At temperatures above the ubiquitin unfolding temperature ($T > 90$ °C), $\underline{U}\underline{B}$ state is the most dominant (eq 16).



Hence, by controlling the temperature, Radley et al.²⁵⁵ controlled distributions of $\underline{U}B$ and $U\underline{B}$ states (eq 17):

$$BU \rightleftharpoons BU \quad (17)$$

Temperature affects all molecules in cells and is not an appropriate “turnkey” in most biological applications. The construct proposed by Radley et al.¹²⁵ is, nevertheless, broadly extendable by replacing the regulating domain, for example ubiquitin, with a ligand-binding domain. Then, one can shift equilibrium in eq 17 to the right alternative species. Such an approach was utilized by the same group.²⁵⁶ Ha et al. used a similar chimeric protein comprising barnase and a ligand-binding polypeptide (GCN4). GCN4 dimerizes via the coiled-coil region at the C-terminal side at nM affinity and binds to DNA. The authors showed that, upon binding to DNA, the chimeric protein results in unfolding of barnase. Hence, by sensing the ligand, the chimeric protein responds with the output—active barnase domain. This study demonstrated a modular biosensor, with the chimeric framework, is potentially widely applicable to many biological problems.

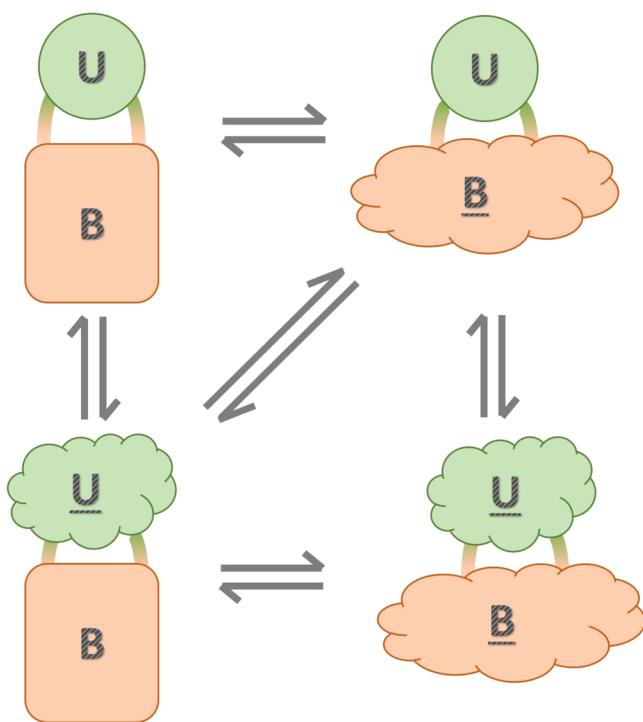


Figure 10. Protein activity regulation through mutually exclusive folding. Ridley et al.¹²⁵ constructed a chimeric protein comprising of fused ubiquitin (U) and barnase (B) proteins so that steric clashes prevent the formation of both folded domains (UB) simultaneously. By using temperature, the authors effectively regulated population distributions between $\underline{U}B$ and $U\underline{B}$ states, where \underline{U} and \underline{B} are the unfolded states of the corresponding proteins.

3.2. Molecular Switch Design by in Vitro Recombination of Nonhomologous Genes

Directed evolution approaches offer a powerful alternative to structure-based approaches for finding ligand-driven molecular switches. Ostermeier's group has developed a platform for designing modular bioswitches using in vitro recombination of nonhomologous genes²⁵⁷ for a series of applications.^{257–260} Guntas et al.^{257,258} created a library of chimeric proteins by randomly inserting a circularly permuted gene encoding TEM1 β -lactamase into the *Escherichia coli* maltose binding protein (MBP). The library was subjected to functional assays whereby the presence of maltose impacts the activity of the TEM1 β -lactamase circular permutant to hydrolyze β -lactam. They identified a reversible bioswitch, RG13, that had increased activity in hydrolyzing β -lactam 25-fold in the presence of maltose and returned to maltose-free activity level upon removal of maltose. Successful creation of an allosteric switch RG13 demonstrated as a proof-of-principle the use of directed evolution for finding regulatable proteins bypassing structure-based identification of the allosteric sites. Guntas and Ostermeier further optimized their constructs and identified chimeric proteins that featured “on–off” behavior with variation of catalytic activity upon maltose binding by 600-fold.²⁵⁹

3.3. Protein Activity Regulation Using Small Molecule Ligands

The original idea to use small molecule ligands to control proteins was proposed by Tucker and Fields.²⁶¹ The authors inserted a small FK506-binding protein (FKBP12; molecular weight, Mw = 12 kDa) into one of the solvent-exposed loops of the dihydrofolate reductase (DHFR) (between residues 107 and 108). The authors transfected this chimeric protein FKBP12-DHFR into a yeast strain lacking endogenous DHFR and measured cell growth of chimeric FKBP12-DHFR. Since yeast cell growth strongly depends on the activity of DHFR, the authors were able to evaluate the activity of transfected constructs. Indeed, the activity of chimeric protein FKBP12-DHFR was shown to strongly depend on the presence of the FK506 ligand in cell viability assays. The authors argued that small molecule ligand FK506 stabilizes the FKBP12 domain of the metastable chimeric protein, and thereby allosterically stabilizes the DHFR active site. To further demonstrate modularity and transferability of their construct, Tucker and Fields²⁶¹ substituted FKBP12 domain by the estrogen receptor- α ligand binding domain and created an (ER α)-DHFR chimeric protein. Similar to the FKBP12-DHFR construct, DHFR activity in the (ER α)-DHFR construct was also regulated by the ER α ligand estrogen. Hence, the authors demonstrated a modular and transferable approach of ligand-driven allosteric control of proteins through insertion of ligand binding domain into the allosteric site of the regulated protein.

The approach of Tucker and Fields²⁶¹ was further developed by Karginov et al.,²⁶² who used focal adhesion kinase (FAK) instead of DHFR, and a small molecule ligand rapamycin instead of FK506 (Figure 11A). Since the termini of FKBP12 are not in proximity, Karginov et al. truncated 22 residues from the N-terminus of FKBP12. This truncated insertable FKBP12 (iFKBP) was inserted into the loop (residues 442–448) flanked by GPG linkers on both sides of iFKBP. Computational studies revealed that iFKBP is significantly destabilized by the truncation of the N-terminal residues, but the binding to rapamycin is preserved: similar to FKBP12, iFKBP dimerizes

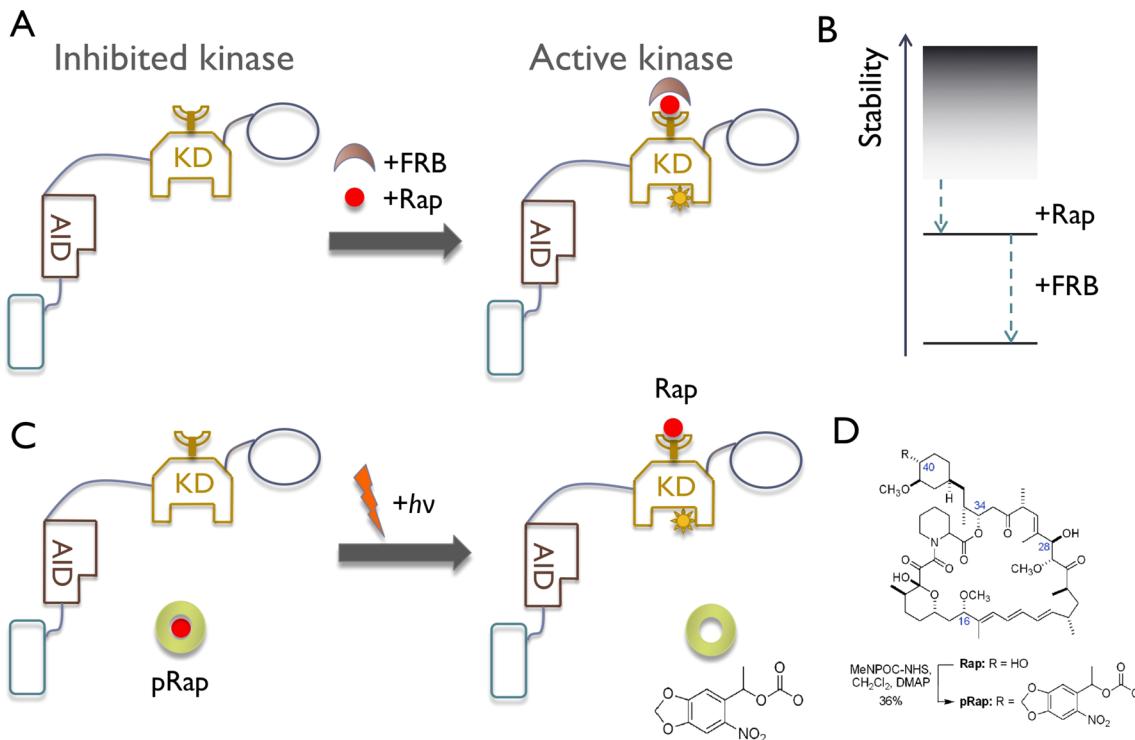


Figure 11. Rapamycin-controlled protein activity. (A) The kinase's catalytic domain (KD) is modified with the inserted iFKBP domain, so that the activity of the KD is eliminated due to allosteric destabilization of the active site. In addition, kinase's autoinhibitory domain (AID) is modified so that it lacks its ability to bind KD, and, hence, inhibit KD. Upon addition of rapamycin Rap, iFKBP is stabilized by the complex formation with FRB facilitated by Rap, resulting in allosteric stabilization of the active site, rendering the kinase active. (B) Free energy diagram of iFKBP only (continuous spectrum), in complex with Rap, and with Rap and FRB. (C) Photoactivatable version of construct (A): caged rapamycin, pRap, is incapable of binding to iFKBP. Membrane-permeable pRap is administered to cells expressing the iFKBP-(kinase) construct. Upon irradiation with the UV light, the caging group leaves, and pure rapamycin remains, which, in turn activates iFKBP-(kinase):FRB complex. (D) Chemical representations of Rap and pRap.

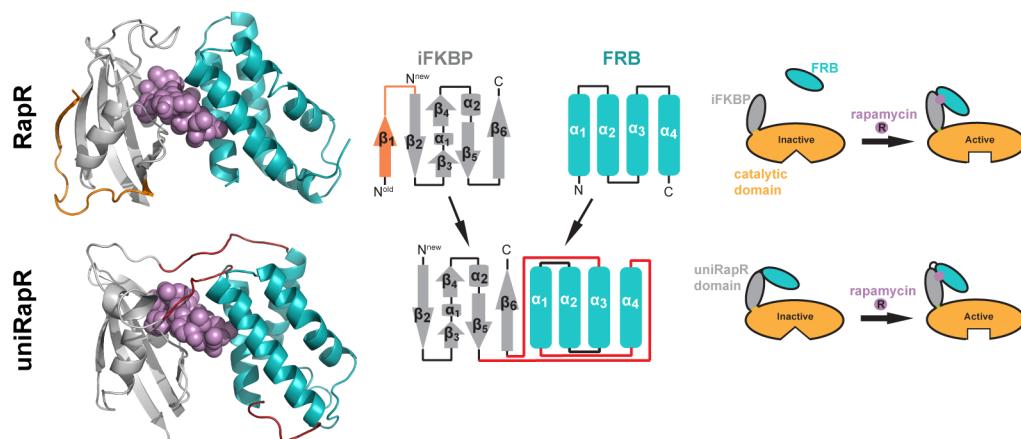


Figure 12. Computationally designed protein domain for protein activity control: contrasting RapR (top panel) and uniRapR (bottom panel) approaches. Left panel: structures of iFKBP:FRB:rapamycin complex and computationally designed structural model of the uniRapR:rapamycin complex. Middle panel: secondary structure representations of iFKBP and FRB (top), and uniRapR (bottom). Right panel: rapamycin-driven activation of iFKBP and FRB (top) and uniRapR (bottom).

with the FKBP12-rapamycin binding (FRB) domain in the presence of rapamycin.²⁶³ Computational analysis revealed strong allosteric coupling between the insertion loop and the functionally critical G-loop of the FAK protein. Further biochemical studies confirmed ligand-dependent phosphorylation of the FAK substrate paxillin at Y31 by the iFKBP-FAK construct. Karginov et al.²⁶⁴ demonstrated the utility of regulatable iFKBP-FAK construct in living cells. iFKBP-FAK

and FRB were transfected into HeLa cells. Cytoskeletal dynamics was monitored using an objective-based total internal reflection fluorescence (TIRF) system. Rapamycin treatment of HeLa cells resulted in formation of large dorsal ruffles, caused by the activation of iFKBP-FAK. Study by Chua et al.²⁶⁵ has demonstrated transferability of this modular system to Src family kinases.

Deiters' lab has extended the construct proposed by Karginov et al.²⁶⁴ to add light regulation onto the iFKBP-(kinase):FRB system by developing caged rapamycin, pRap, which is incapable of forming iFKBP:FRB complex unless it is irradiated with UV light, thereby, producing rapamycin and forming iFKBP:FRB complex²⁶² (Figure 11B). Deiters' lab synthesized pRap from rapamycin by selective acylation with the mixed carbonate of *R*-methyl-6-nitropiperonyl alcohol and *N*-hydroxysuccinimide (MeN-POC-NHS).²⁶² In a similar fashion, Brown et al.²⁶⁶ developed the light activated rapamycin dimer dRap that does not induce iFKBP:FRB complex formation until exposed to light. This work demonstrated the utility of their new light-activatable dimerizer on a range of biologically relevant systems.

One drawback of the iFKBP-(kinase) system is that it relies on transfection of this chimeric construct and the FRB protein, making direct interrogations of biological systems complex. To circumvent this pitfall, Dagliyan at al.²⁶⁷ computationally designed a protein uniRapR that featured iFKBP:FRB binding site. uniRapR was thermodynamically unstable without rapamycin and stable in the presence of ligand (Figure 12). The kinase insertion site remained unaltered from the previous studies.²⁶⁴ In biochemical studies Dagliyan at al.²⁶⁷ demonstrated that the uniRapR-Src chimeric protein phosphorylates paxillin at Y31 in a rapamycin-dependent manner. In living (HeLa) cells activation of uniRapR-Src with rapamycin led to marked cell area increase and polarized spreading, highlighting the role of Src in cancer metastasis. The authors further demonstrated transferability of the developed system to other kinases, such as p38 and FAK. By engineering the transgenic zebrafish, Dagliyan at al.²⁶⁷ demonstrated rapamycin-dependent activation of Src, which induces loss of cell–cell contacts in epidermis, a hallmark of the epithelial–mesenchymal transition. This study was important because (i) for the first time two states of a protein were designed using computational approaches,^{268–270} (ii) it demonstrated transferability of the modular design to other kinases, and (iii) it showed the usefulness of this design not only in cellular studies, but also in a living organism.

3.4. Inducing Large Conformational Changes Using Metals

The relationship between protein sequence and structure is nontrivial. Proteins sharing at least 20% of sequence similarity are likely to share similar structures.^{54,271–277} Yet, examples exist in nature of two proteins belonging to different fold families that differ by only a single amino acid or a few amino acids. These proteins are called chameleon proteins,^{278–281} and they have been argued to be “bridges” between various fold families in the course of evolution.²⁸² From a protein design perspective, chameleon proteins present a potentially tractable target for developing a protein switch, which can undergo a conformational transition from one state to another.^{283–294} Ambroggio and Kuhlman proposed an innovative computational approach to design peptides that switch between a helical bundle and a Zn-finger fold.²⁹⁵ The templates of the corresponding states were an α -helical peptide of hemagglutinin²⁹⁶ (residues 13–44, PDB ID 1HTM) that forms a three helix bundle (Figure 13A) and a zinc finger Zif268²⁹⁷ (residues 3–33, PDB ID 1AAY). Ambroggio and Kuhlman then structurally aligned these templates and performed computational protein design using RosettaDesign.^{101,298} To control the switching between these two states, a metal-binding site in the zinc finger remained completely fixed during computational

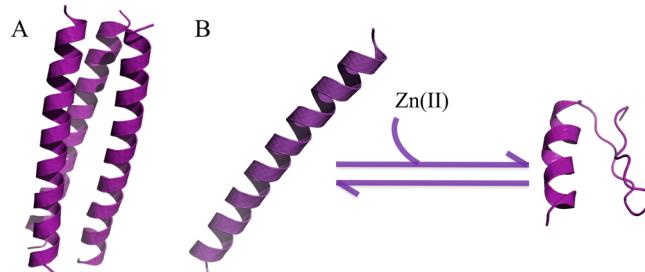


Figure 13. Controlling allostery with metals. (A) Structure of the three helix bundle formed by the α -helical peptide of hemagglutinin²⁹⁶ (residues 13–44, PDB ID 1HTM). (B) Redesigned peptide, sw2, which undergoes conformational change upon the binding of zinc, mimicking the zinc finger Zif268²⁹⁷ (residues 3–33, PDB ID 1AAY).

design simulations. The authors noticed that, in protein switches, amino acids typically strongly favor one or another conformation. This conformational duality is potentially responsible for bistable switching.²⁹⁹ Hence, they modified the RosettaDesign scoring function to emphasize this duality by promoting amino acids that strongly favor coiled-coil and zinc fingers:

$$E_{\text{MSA}}(\text{aa}, i) = -\ln\left(\frac{1}{2}[P_{\text{cc}}(\text{aa}, i) + P_{\text{ZnF}}(\text{aa}, i)]\right) \quad (18)$$

where $P_{\text{cc}}(\text{aa}, i)$ and $P_{\text{ZnF}}(\text{aa}, i)$ are the propensities of an amino acid aa at the position i to form a coiled-coil or a zinc finger, correspondingly. The modified energy function $E_{\text{MSA}}(\text{aa}, i)$ was derived from multiple sequence alignments during the course of design. In biophysical assays, the lowest energy design, sw2, exhibited a strong preference for the helical state in the absence of Zn(II) and for the zinc finger state in the presence of Zn(II) (Figure 13B). Furthermore, the authors demonstrated the ability of this protein to bind cobalt in its metal-binding site, which has the same coordination as zinc, offering further support for the zinc finger state.

A more ad hoc design was performed by Cerasoli et al.²⁹⁴ the zinc-binding/coiled coil (ZiCo) peptide is based on merging hemagglutinin and zinc finger structures. The ZiCo peptide exhibited properties similar to those proposed by Ambroggio and Kuhlman. These studies demonstrate the feasibility of inducing controlled conformational changes.³⁰⁰ The ability to control large conformational rearrangements of proteins using metal ions alone provides an appealing avenue for designing allosteric modulators and sensors.

3.5. Regulation through Engineered Allostery

3.5.1. Grafting Allostery from One Orthologous Protein to Another. Understanding molecular evolution has a direct impact on our ability to engineer new molecules and the activities associated with them. Cross et al. has recently³⁰¹ attempted to understand how protein regulation evolved across orthologous proteins. The authors studied 3-deoxy-D-arabino-heptulosonate 7-phosphate synthases (DAH7PS), which, depending on the organism, do (e.g., *Thermotoga maritima*) or do not (e.g., *Pyrococcus furiosus*) possess a regulatory ACT domain.³⁰² The authors fused two orthologous proteins: regulatory ACT domain *T. maritima* and the catalytic domain DAH7PS from *P. furiosus*. The chimeric protein *T^{ma}*ACT-*P^{fu}*DAH7PS featured allosteric regulation present in *T. maritima* but not in *P. furiosus*. The authors demonstrated

that one could introduce allosteric regulation through grafting regulatory domains in distantly evolutionary related proteins.

3.5.2. Engineering Allosteric Regulation Using Next-Generation Sequencing. Recent technological advances in next-generation sequencing and DNA synthesis allowed Church's group to develop a framework for designing allosteric regulators.³⁰³ They used LacI repressor,³⁰⁴ which regulates the lactose-utilization operon *lac*, from *E. coli* as a model system and a deep mutational scanning technique developed by Fowler and Fields.³⁰⁵ Raman et al.³⁰³ engineered an orthogonal transcription factor by simultaneously redesigning the ligand-binding domain and the DNA-binding domain. The authors further outlined strategies for developing other allosterically controlled proteins, such as direct/indirect transcriptional readout, split reporter assay, environmentally sensitive fluorophores, and domain-insertion reporters.

4. ALLOSTERIC CONTROL OF PROTEINS WITH LIGHT

Controlling proteins with light presents an outstanding opportunity for rapid interrogation of cellular life. Light control of engineered proteins, while less applicable to tissue and organism scale studies, is by far faster than ligand control of proteins: ligand control is limited by the diffusion of the ligand and the binding affinity of the regulated protein. Next, we describe a pioneering approach for controlling proteins using photoactivation, whereby the target or host protein is activated in the lit state, while inactive in the dark state.

Strickland et al.³⁰⁶ proposed a construct that utilizes light induced allosteric switching to control trp repressor from *E. coli* by photoreactive LOV2 domain from *Avena sativa* protein phototrophin 1.³⁰⁷ They utilized properties of the LOV2 domain and transcription factor trp repressor (TrpR) protein to contain a helical secondary structure on the opposite C-terminal ($\text{J}\alpha$) and N-terminal of the corresponding proteins. They fused these helices and truncated a part of the trp receptor helix (Figure 14). The length of the helix was chosen

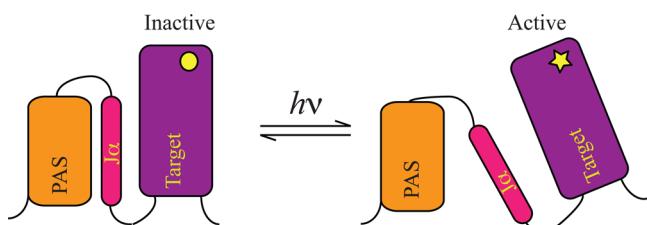


Figure 14. Photoactivation. Chimeric protein constructed from the fusion of the target possessing helical N-terminal with the LOV2 domain's C-terminal $\text{J}\alpha$ helix. In the dark state the target is allosterically inhibited (blue path). Light triggers $\text{J}\alpha$ dissociation from PAS domain, rendering the target protein active.

such that it sterically interferes with either the LOV2 domain or TrpR, making the construct bistable. The structure fluctuated between a state bound to the LOV2 domain and a state bound to TrpR. Upon photoactivation, a photon triggers covalent cross-linking between a conserved cysteine and a flavin mononucleotide, which then unfurls $\text{J}\alpha$ from the LOV2 domain. By penalizing the state where $\text{J}\alpha$ is bound to the LOV2 domain, the state where the extended $\text{J}\alpha$ helix is bound to TrpR becomes dominant. In this state, homodimeric TrpR binds its operator DNA. The authors used multiple constructs (12), but found that only one, LovTAP, preferentially protects cognate DNA from nuclease digestion.

Strickland's et al.³⁰⁶ construct is modular as it connects termini of the host protein and the photoactivatable protein LOV2. However, there are two potential drawbacks of this modular design. First, the construct relies on the host protein to form an α helix at the N-terminal. Second, the N-terminal of the host protein may not be allosterically connected to the site of interest, e.g. an active site or a binding site. These limitations decrease the range of applicability of the design proposed by Strickland et al.

A similar approach was also undertaken by Cosentino et al.,³⁰⁸ who engineered blue-light-induced K^+ channel 1 (BLINK1) by fusing LOV2 domain to the N-terminus of the small viral K^+ channel Kcv. Photoactivation of BLINK1 resulted in reversible opening of the K^+ channel, which hyperpolarized the cell to K^+ equilibrium potential. The authors further demonstrated the utility of BLINK1 regulator in zebrafish larvae, which supports broad applicability of LOV2-based constructs in optogenetic applications.

Ranganathan's group used their SCA approach for finding allosteric sites (see section 2.3.2) to identify surface residues of the target protein, *E. coli* dihydrofolate reductase (DHFR), that are allosterically coupled with the active site.³⁰⁹ They used an alignment of DHFR 418 family members and identified the $\beta\text{F}-\beta\text{G}$ surface-exposed loop that featured strong evolutionary coupling to the distant enzyme active site. They inferred allosteric coupling based on the SCA analysis and made several LOV2-DHFR constructs, whereby the LOV2 N- and C-termini were inserted in the $\beta\text{F}-\beta\text{G}$ loop. They showed that upon light activation one of the LOV2-DHFR chimeras responded with 1.6-fold overall increase of the turnover rate k_{cat} at 25 °C, and 2-fold increase at 17 °C. Lee et al.³⁰⁹ also showed that a control insertion site that did not feature evolutionary coupling, as detected by their SCA method, did not show any alteration in DHFR activity between lit and dark states. Despite the modest light dependent increase and significantly lower activity with respect to that of wild type, the construct has showed as a proof-of-principle an approach for allosteric photoactivation of an enzyme.

5. CONTROLLING PROTEIN DYNAMICS AND STABILITY BY TARGETING ALLOSTERIC NETWORKS

5.1. Allosteric Rescue of CFTR

Cystic fibrosis³¹⁰ is a genetic disease caused by mutations in CFTR, an epithelial chloride ion channel present in airway, intestine, pancreas, and urogenital systems.^{311–313} Deletion of F508 in nucleotide binding domain 1 (NBD1) has been associated with approximately 90% of cystic fibrosis disease cases. Numerous studies have demonstrated that ΔF508 has a dramatic destabilizing effect on CFTR structure and dynamics.^{314,315} While targeting the F508 site with small molecules has been considered by a number of scientists, no viable drugs have been proposed to target the F508 site and rescue CFTR. As an allosteric rescue may potentially become a viable strategy to restore CFTR stability and function, it became critical to detect allosteric couplings between the F508 site and other surface exposed residues and corresponding allosteric pathways.

Using covariation analysis and discrete molecular dynamics simulations^{134,135,316} of NBD1, Aleksandrov and colleagues were able to detect long-distance coupling in the CFTR NBD1 domain, specifically between mutant site F508 and a 32-residue segment region, regulatory insertion (RI, residues 404–435), whose dynamics indirectly influence tertiary interactions

between NBD1 and the rest of the protein by disrupting the ATP-binding site and F508 site.¹⁴⁷ The authors validated this connection by removing the regulatory insertion region to decrease allosterically disruptive fluctuations, and observed complete restoration of CFTR-ΔF508 function to a level comparable to the wild type protein.

Proctor and colleagues applied optimal path mapping algorithm to uncover allosteric networks that connect CFTR disease-associated mutant sites F508 and I507 to other surface residues on NBD1.¹¹⁶ The authors determined that, in the CFTR-ΔF508 mutant, the loop containing the F508 site, the structurally diverse region (SDR, residues 532–552), and the RI-SDR bridge (residues 492–502), RI, and part of the ATP-binding subdomain (residues 570–600) are mutually connected, such that perturbations in one of these regions is reflected on others. In agreement with previous findings, the deletion of F508 results in increased fluctuations of the RI region, which ultimately leads to higher mobility of the ATP-binding region and destabilization of CFTR. In contrast, in the CFTR-ΔI507 mutant, the mutation results in redistribution of the network fluctuations to neighboring regions, resulting in non-native coupling between the SDR and the ATP-binding subdomain. Proctor et al.¹¹⁶ further identified critical residues responsible for a coupled fluctuation network, removal of which would disrupt the signal propagation throughout the network. In CFTR-ΔF508 mutant, residue S492 was found to have the largest effect on the network. Through computational redesign,^{269,270} S492P mutation was found to disrupt the network in NBD1. Experimental testing of S492P mutant confirmed complete restoration of CFTR-ΔF508 function, thereby validating the predictions.³¹⁷ In CFTR-ΔI507 mutant, critical nodes (bottlenecks) reside in regions essential for CFTR maturation and function and, hence, are not amenable for rescue. It was found that S492P mutant results in only partial rescue of CFTR-ΔI507. Interestingly, from the sequence alignment of NBD1 and NBD2 domains from various members of ABC proteins, Ser and Pro were exclusively found at the position 492. Notably, in higher body temperature animals, like chickens, CFTR features Pro at the position 492, suggesting evolutionary selection for thermally stabilizing amino acid at this site.

6. OUTLOOK

Allostery in proteins has been a topic of interest in the scientific community for over half a century due its central role in many biological processes. Engineering allostery to control protein function and, therefore, cellular phenotypes is a relatively recent (decade-old) endeavor brought by the discovery of light-sensitive proteins, recent advances in directed evolution, high-throughput screening, deep mutational scanning techniques, and computational protein design. These approaches offered unprecedented control of proteins and created venues for complex interrogations of biological systems. The wealth of alternative approaches for allosteric control of proteins open doors for examination of cascades of cellular processes at the systems level by combining (multiplexing) these approaches: simultaneous regulation of multiple proteins can be utilized for understanding signal routing in living cells.

While most of the developed techniques to control molecular activities have been applied to proteins, RNA represents the next frontier for allosteric regulation. Unlike proteins, RNA has not been explored either structurally or functionally, yet we are beginning to understand how critical, diverse, and multifaceted

are the roles of RNA in cellular life.^{318–322} From the structural perspectives, a much smaller number of RNA structures have been solved (~1000) compared to proteins (>100 000). Experimental and computational difficulties hinder RNA structure determination efforts, although new technological advances^{323–329} offer promise for better understanding RNA structure, dynamics, and allosteric regulation. Since RNA folding processes (diffusion–collision³³⁰) are conceptually different from those of proteins (nucleation^{75–77,82–88}), it is unclear whether the technologies developed for detecting allostery in proteins are transferable to that in RNA. Hence, novel approaches to control RNA molecules may offer unorthodox means for interrogation of cellular life.

AUTHOR INFORMATION

Corresponding Author

*E-mail: dokh@unc.edu.

Notes

The authors declare no competing financial interest.

Biography

Dr. Dokholyan received his Ph.D. in physics in 1999 at Boston University under the supervision of Dr. H. Eugene Stanley. His graduate work spanned the area of statistical mechanics and its applications in biological macromolecules. Upon graduation, Dr. Dokholyan joined Dr. Eugene Shakhnovich at Harvard University in the Department of Chemistry and Chemical Biology as an NIH NRSA Fellow. His work was focused on protein folding, design, and evolution. Dr. Dokholyan joined the Department of Biochemistry and Biophysics at the University of North Carolina at Chapel Hill School of Medicine as an assistant professor in 2002, was promoted to associate professor in 2008, and was promoted to full professor in 2011. Dr. Dokholyan has served as the Director of the Center for Computational and Systems Biology and the Graduate Director of the Program in Molecular and Cellular Biophysics at UNC. Dr. Dokholyan has published over 200 peer-reviewed articles and 17 book chapters. Dr. Dokholyan was named a Fellow of the American Physical Society in 2013. In 2014, Dr. Dokholyan was named the Michael Hooker Distinguished Professor of Biochemistry and Biophysics. Over the past decade, Dr. Dokholyan's work has followed two principal directions: (i) development of methodologies to understand and engineer molecular structure and function, and (ii) application of these methodologies to understand the etiologies of human diseases toward the development of therapeutic strategies. Dr. Dokholyan has developed computational platforms for modeling and design of molecular structure and drug discovery, and has made critical strides in elucidating the molecular mechanisms of amyotrophic lateral sclerosis and cystic fibrosis.

ACKNOWLEDGMENTS

We acknowledge the support of the National Institutes of Health Grants SR01GM080742 and R01AI102732. We also thank Marino Convertino, Onur Dagliyan, Aspen Gutgsell, Andrey Krokhotin, David Mowrey, Konstantin Popov, Elizabeth A. Proctor, and Benfeard Williams for their help with this review.

REFERENCES

- (1) Monod, J.; Changeux, J.; Jacob, F. Allosteric Proteins and Cellular Control Systems. *J. Mol. Biol.* **1963**, *6*, 306–329.
- (2) Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **1958**, *44*, 98–104.

- (3) Frauenfelder, H.; McMahon, B. H.; Austin, R. H.; Chu, K.; Groves, J. T. The Role of Structure, Energy Landscape, Dynamics, and Allostery in the Enzymatic Function of Myoglobin. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 2370–2374.
- (4) Adams, J.; Johnson, K.; Matthews, R.; Benkovic, S. J. Effects of Distal Point-site Mutations on the Binding and Catalysis of Dihydrofolate Reductase from Escherichia Coli. *Biochemistry* **1989**, *28*, 6611–6618.
- (5) Green, S. M.; Shortle, D. Patterns of Nonadditivity Between Pairs of Stability Mutations in Staphylococcal Nuclease. *Biochemistry* **1993**, *32*, 10131–10139.
- (6) Mace, J. E.; Wilk, B. J.; Agard, D. A. Functional Linkage Between the Active Site of Alpha-lytic Protease and Distant Regions of Structure: Scanning Alanine Mutagenesis of a Surface Loop Affects Activity and Substrate Specificity. *J. Mol. Biol.* **1995**, *251*, 116–134.
- (7) Olsen, D. B.; Stahlhut, M. W.; Rutkowski, C. A.; Schock, H. B.; vanOlden, A. L.; Kuo, L. C. Non-active Site Changes Elicit Broad-based Cross-resistance of the HIV-1 Protease to Inhibitors. *J. Biol. Chem.* **1999**, *274*, 23699–23701.
- (8) Meroueh, S. O.; Roblin, P.; Golemi, D.; Maveyraud, L.; Vakulenko, S. B.; Zhang, Y.; Samama, J. P.; Mobashery, S. Molecular Dynamics at the Root of Expansion of Function in the M69L Inhibitor-resistant TEM Beta-lactamase from Escherichia Coli. *J. Am. Chem. Soc.* **2002**, *124*, 9422–9430.
- (9) Duke, T. A.; Le Novère, N.; Bray, D. Conformational Spread in a Ring of Proteins: a Stochastic Approach to Allostery. *J. Mol. Biol.* **2001**, *308*, 541–553.
- (10) Pasqualato, S.; Ménétrey, J.; Franco, M.; Cherfils, J. The Structural GDP/GTP Cycle of Human Arf6. *EMBO Rep.* **2001**, *2*, 234–238.
- (11) Soisson, S. M.; MacDougall-Shackleton, B.; Schleif, R.; Wolberger, C. Structural Basis for Ligand-regulated Oligomerization of AraC. *Science* **1997**, *276*, 421–425.
- (12) Strater, N.; Hakansson, K.; Schnappauf, G.; Braus, G.; Lipscomb, W. N. Crystal Structure of the T State of Allosteric Yeast Chorismate Mutase and Comparison with the R State. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 3330–3334.
- (13) Oliva, G.; Fontes, M. R.; Garratt, R. C.; Altamirano, M. M.; Calcagno, M. L.; Horjales, E. Structure and Catalytic Mechanism of Glucosamine 6-phosphate Deaminase from Escherichia Coli at 2.1 Å Resolution. *Structure* **1995**, *3*, 1323–1332.
- (14) Wiesmann, C.; Barr, K. J.; Kung, J.; Zhu, J.; Erlanson, D. A.; Shen, W.; Fahr, B. J.; Zhong, M.; Taylor, L.; Randal, M.; McDowell, R. S.; Hansen, S. K. Allosteric Inhibition of Protein Tyrosine Phosphatase 1B. *Nat. Struct. Mol. Biol.* **2004**, *11*, 730–737.
- (15) Makhlynets, O. V.; Raymond, E. A.; Korendovych, I. V. Design of Allosterically Regulated Protein Catalysts. *Biochemistry* **2015**, *54*, 1444–1456.
- (16) Pauling, L. The Oxygen Equilibrium of Hemoglobin and Its Structural Interpretation. *Proc. Natl. Acad. Sci. U. S. A.* **1935**, *21*, 186–191.
- (17) Gerhart, J. C.; Pardee, A. B. The Enzymology of Control by Feedback Inhibition. *J. Biol. Chem.* **1962**, *237*, 891–896.
- (18) Stadtman, E. R.; Cohen, G. N.; Lebras, G. Feedback Inhibition and Repression of Aspartokinase Activity in Escherichia Coli. *Ann. N. Y. Acad. Sci.* **1961**, *94*, 952–959.
- (19) Patte, J. C.; Le Bras, G.; Loviny, T.; Cohen, G. N. Retro-inhibition and Repression of the Homoserine Dehydrogenase of Escherichia Coli. *Biochim. Biophys. Acta, Spec. Sect. Enzymol. Subj.* **1963**, *67*, 16–30.
- (20) Monod, J.; Wyman, J.; Changeux, J. P. On The Nature Of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* **1965**, *12*, 88–118.
- (21) Cooper, A.; Dryden, D. T. Allostery Without Conformational Change. A Plausible Model. *Eur. Biophys. J.* **1984**, *11*, 103–109.
- (22) Ptitsyn, O. B. Inter-domain Mobility in Proteins and Its Probable Functional Role. *FEBS Lett.* **1978**, *93*, 1–4.
- (23) Birktoft, J. J.; Kraut, J.; Freer, S. T. A Detailed Structural Comparison Between the Charge Relay System in Chymotrypsinogen and in Alpha-chymotrypsin. *Biochemistry* **1976**, *15*, 4481–4485.
- (24) Yonath, A.; Podjarny, A.; Honig, B.; Sielecki, A.; Traub, W. Crystallographic Studies of Protein Denaturation and Renaturation. 2. Sodium Dodecyl Sulfate Induced Structural Changes in Triclinic Lysozyme. *Biochemistry* **1977**, *16*, 1418–1424.
- (25) Levitt, M.; Warshel, A. Computer Simulation of Protein Folding. *Nature* **1975**, *253*, 694–698.
- (26) McCammon, J. A.; Gelin, B. R.; Karplus, M.; Wolynes, P. G. The Hinge-bending Mode in Lysozyme. *Nature* **1976**, *262*, 325–326.
- (27) Honig, B.; Ray, A.; Levinthal, C. Conformational Flexibility and Protein Folding: Rigid Structural Fragments Connected by Flexible Joints in Subtilisin BPN. *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73*, 1974–1978.
- (28) Gunasekaran, K.; Ma, B.; Nussinov, R. Is Allostery an Intrinsic Property of All Dynamic Proteins? *Proteins: Struct., Funct., Genet.* **2004**, *57*, 433–443.
- (29) Kar, G.; Keskin, O.; Gursoy, A.; Nussinov, R. Allostery and Population Shift in Drug Discovery. *Curr. Opin. Pharmacol.* **2010**, *10*, 715–722.
- (30) Zhuravlev, P. I.; Papoian, G. A. Protein Functional Landscapes, Dynamics, Allostery: a Tortuous Path Towards a Universal Theoretical Framework. *Q. Rev. Biophys.* **2010**, *43*, 295–332.
- (31) Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N. Structural and Energetic Basis of Allostery. *Annu. Rev. Biophys.* **2012**, *41*, 585–609.
- (32) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. The Ensemble Nature of Allostery. *Nature* **2014**, *508*, 331–339.
- (33) Volkman, B. F.; Lipson, D.; Wemmer, D. E.; Kern, D. Two-state Allosteric Behavior in a Single-domain Signaling Protein. *Science* **2001**, *291*, 2429–2433.
- (34) Popovych, N.; Sun, S.; Ebright, R. H.; Kalodimos, C. G. Dynamically Driven Protein Allostery. *Nat. Struct. Mol. Biol.* **2006**, *13*, 831–838.
- (35) Gsponer, J.; Christodoulou, J.; Cavalli, A.; Bui, J. M.; Richter, B.; Dobson, C. M.; Vendruscolo, M. A Coupled Equilibrium Shift Mechanism in Calmodulin-mediated Signal Transduction. *Structure* **2008**, *16*, 736–746.
- (36) Tzeng, S. R.; Kalodimos, C. G. Protein Dynamics and Allostery: An NMR View. *Curr. Opin. Struct. Biol.* **2011**, *21*, 62–67.
- (37) Tzeng, S. R.; Kalodimos, C. G. Dynamic Activation of an Allosteric Regulatory Protein. *Nature* **2009**, *462*, 368–372.
- (38) Boehr, D. D.; Nussinov, R.; Wright, P. E. The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. *Nat. Chem. Biol.* **2009**, *5*, 789–796.
- (39) Frantom, P. A.; Zhang, H. M.; Emmett, M. R.; Marshall, A. G.; Blanchard, J. S. Mapping of the Allosteric Network in the Regulation of Alpha-isopropylmalate Synthase from Mycobacterium Tuberculosis by the Feedback Inhibitor L-leucine: Solution-phase H/D Exchange Monitored by FT-ICR Mass Spectrometry. *Biochemistry* **2009**, *48*, 7457–7464.
- (40) Cui, Q.; Karplus, M. Allostery and Cooperativity Revisited. *Protein Sci.* **2008**, *17*, 1295–1307.
- (41) Kim, S.; Broströmer, E.; Xing, D.; Jin, J.; Chong, S.; Ge, H.; Wang, S.; Gu, C.; Yang, L.; Gao, Y. Q.; Su, X. D.; Sun, Y.; Xie, X. S. Probing Allostery through DNA. *Science* **2013**, *339*, 816–819.
- (42) Vitreschak, A. G.; Rodionov, D. A.; Mironov, A. A.; Gelfand, M. S. Riboswitches: The Oldest Mechanism for the Regulation of Gene Expression? *Trends Genet.* **2004**, *20*, 44–50.
- (43) Nudler, E.; Mironov, A. S. The Riboswitch Control of Bacterial Metabolism. *Trends Biochem. Sci.* **2004**, *29*, 11–17.
- (44) Henkin, T. M. Riboswitch RNAs: Using RNA to Sense Cellular Metabolism. *Genes Dev.* **2008**, *22*, 3383–3390.
- (45) Ren, A.; Rajashankar, K. R.; Patel, D. J. Fluoride Ion Encapsulation by Mg²⁺ Ions and Phosphates in a Fluoride Riboswitch. *Nature* **2012**, *486*, 85–89.
- (46) Breaker, R. R. Riboswitches and the RNA World. *Cold Spring Harbor Perspect. Biol.* **2012**, *4*, a003566.

- (47) Baker, J. L.; Sudarsan, N.; Weinberg, Z.; Roth, A.; Stockbridge, R. B.; Breaker, R. R. Widespread Genetic Switches and Toxicity Resistance Proteins for Fluoride. *Science* **2012**, *335*, 233–235.
- (48) Tsai, C. J.; Nussinov, R. A Unified View of “How Allostery Works.” *PLoS Comput. Biol.* **2014**, *10*, e1003394.
- (49) Christopoulos, A. Allosteric Binding Sites on Cell-surface Receptors: Novel Targets for Drug Discovery. *Nat. Rev. Drug Discovery* **2002**, *1*, 198–210.
- (50) Lindsley, C. W. 2013 Philip S. Portoghesi Medicinal Chemistry Lectureship: Drug Discovery Targeting Allosteric Sites. *J. Med. Chem.* **2014**, *57*, 7485–7498.
- (51) Wentur, C. J.; Gentry, P. R.; Mathews, T. P.; Lindsley, C. W. Drugs for Allosteric Sites on Receptors. *Annu. Rev. Pharmacol. Toxicol.* **2014**, *54*, 165–184.
- (52) Szilágyi, A.; Nussinov, R.; Csermely, P. Allo-network Drugs: Extension of the Allosteric Drug Concept to Protein-Protein Interaction and Signaling Networks. *Curr. Top. Med. Chem.* **2013**, *13*, 64–77.
- (53) Nussinov, R.; Tsai, C. J.; Csermely, P. Allo-network Drugs: Harnessing Allostery in Cellular Networks. *Trends Pharmacol. Sci.* **2011**, *32*, 686–693.
- (54) Dokholyan, N. V.; Shakhnovich, E. I. Understanding Hierarchical Protein Evolution from First Principles. *J. Mol. Biol.* **2001**, *312*, 289–307.
- (55) Ramachandran, S.; Vogel, L.; Strahl, B. D.; Dokholyan, N. V. Thermodynamic Stability of Histone H3 Is a Necessary but Not Sufficient Driving Force for Its Evolutionary Conservation. *PLoS Comput. Biol.* **2011**, *7*, e1001042.
- (56) Proctor, E. A.; Kota, P.; Demarest, S. J.; Caravella, J. A.; Dokholyan, N. V. Highly Covarying Residues Have a Functional Role in Antibody Constant Domains. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 884–895.
- (57) Kannan, N.; Wu, J.; Anand, G. S.; Yooseph, S.; Neuwald, A. F.; Venter, J. C.; Taylor, S. S. Evolution of Allostery in the Cyclic Nucleotide Binding Module. *Genome Biol.* **2007**, *8*, R264.
- (58) Shukla, D.; Meng, Y.; Roux, B.; Pande, V. S. Activation Pathway of Src Kinase Reveals Intermediate States as Targets for Drug Design. *Nat. Commun.* **2014**, *5*, 3397.
- (59) Kohl, A.; Amstutz, P.; Parizek, P.; Binz, H. K.; Briand, C.; Capitani, G.; Forrer, P.; Plückthun, A.; Grütter, M. G. Allosteric Inhibition of Aminoglycoside Phosphotransferase by a Designed Ankyrin Repeat Protein. *Structure* **2005**, *13*, 1131–1141.
- (60) Yang, Z.; Májek, P.; Bahar, I. Allosteric Transitions of Supramolecular Systems Explored by Network Models: Application to Chaperonin GroEL. *PLoS Comput. Biol.* **2009**, *5*, e1000360.
- (61) Csermely, P.; Palotai, R.; Nussinov, R. Induced Fit, Conformational Selection and Independent Dynamic Segments: An Extended View of Binding Events. *Trends Biochem. Sci.* **2010**, *35*, 539–546.
- (62) Novinec, M.; Korenč, M.; Caflisch, A.; Ranganathan, R.; Lenarčič, B.; Baici, A. A Novel Allosteric Mechanism in the Cysteine Peptidase Cathepsin K Discovered by Computational Methods. *Nat. Commun.* **2014**, *5*, 3287.
- (63) Vanatta, D. K.; Shukla, D.; Lawrenz, M.; Pande, V. S. A Network of Molecular Switches Controls the Activation of the Two-component Response Regulator NtrC. *Nat. Commun.* **2015**, *6*, 7283.
- (64) Sumbul, F.; Acuner-Ozbabacan, S. E.; Haliloglu, T. Allosteric Dynamic Control of Binding. *Biophys. J.* **2015**, *109*, 1190–1201.
- (65) Zhuravleva, A.; Gerasch, L. M. Allosteric Signal Transmission in the Nucleotide-binding Domain of 70-kDa Heat Shock Protein (Hsp70) Molecular Chaperones. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 6987–6992.
- (66) Selvaratnam, R.; Chowdhury, S.; VanSchouwen, B.; Melacini, G. Mapping Allostery through the Covariance Analysis of NMR Chemical Shifts. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 6133–6138.
- (67) Buchli, B.; Waldauer, S. A.; Walser, R.; Denten, M. L.; Pfister, R.; Blöchliger, N.; Steiner, S.; Caflisch, A.; Zerbe, O.; Hamm, P. Kinetic Response of a Photoperturbed Allosteric Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 11725–11730.
- (68) Henry, E. R.; Mozzarelli, A.; Viappiani, C.; Abbruzzetti, S.; Bettati, S.; Ronda, L.; Bruno, S.; Eaton, W. A. Experiments on Hemoglobin in Single Crystals and Silica Gels Distinguish Among Allosteric Models. *Biophys. J.* **2015**, *109*, 1264–1272.
- (69) Lockless, S. W.; Ranganathan, R. Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science* **1999**, *286*, 295–299.
- (70) Watts, D. J.; Strogatz, S. H. Collective Dynamics of “Small-world” Networks. *Nature* **1998**, *393*, 440–442.
- (71) Dokholyan, N. V.; Li, L.; Ding, F.; Shakhnovich, E. I. Topological Determinants of Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 8637–8641.
- (72) Vendruscolo, M.; Dokholyan, N.; Paci, E.; Karplus, M. Small-world View of the Amino Acids That Play a Key Role in Protein Folding. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2002**, *65*, 061910.
- (73) Rose, P. W.; Prlić, A.; Bi, C.; Bluhm, W. F.; Christie, C. H.; Dutta, S.; Green, R. K.; Goodsell, D. S.; Westbrook, J. D.; Woo, J.; Young, J.; Zardecki, C.; Berman, H. M.; Bourne, P. E.; Burley, S. K. The RCSB Protein Data Bank: Views of Structural Biology for Basic and Applied Research and Education. *Nucleic Acids Res.* **2015**, *43*, D345–D356.
- (74) Ding, F.; Jha, R. K.; Dokholyan, N. V. Scaling Behavior and Structure of Denatured Proteins. *Structure* **2005**, *13*, 1047–1054.
- (75) Fersht, A. R. Nucleation Mechanisms in Protein Folding. *Curr. Opin. Struct. Biol.* **1997**, *7*, 3–9.
- (76) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and Energetic Factors: What Determines the Structural Details of the Transition State Ensemble and “En-route” Intermediates for Protein Folding? An Investigation for Small Globular Proteins. *J. Mol. Biol.* **2000**, *298*, 937–953.
- (77) Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. Identifying the Protein Folding Nucleus Using Molecular Dynamics. *J. Mol. Biol.* **2000**, *296*, 1183–1188.
- (78) Ding, F.; Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. Direct Molecular Dynamics Observation of Protein Folding Transition State Ensemble. *Biophys. J.* **2002**, *83*, 3525–3532.
- (79) Dokholyan, N. V. Studies of Folding and Misfolding Using Simplified Models. *Curr. Opin. Struct. Biol.* **2006**, *16*, 79–85.
- (80) Dokholyan, N. V.; Shakhnovich, B.; Shakhnovich, E. I. Expanding Protein Universe and Its Origin from the Biological Big Bang. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 14132–14136.
- (81) Dokholyan, N. V. What Is the Protein Design Alphabet? *Proteins: Struct., Funct., Genet.* **2004**, *54*, 622–628.
- (82) Shakhnovich, E. I.; Finkelstein, A. V. Theory of Cooperative Transitions in Protein Molecules. I. Why Denaturation of Globular Protein Is a First-order Phase Transition. *Biopolymers* **1989**, *28*, 1667–1680.
- (83) Jackson, S. E.; Fersht, A. R. Folding of Chymotrypsin Inhibitor 2. 1. Evidence for a Two-state Transition. *Biochemistry* **1991**, *30*, 10428–10435.
- (84) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. Specific Nucleus as the Transition State for Protein Folding: Evidence from the Lattice Model. *Biochemistry* **1994**, *33*, 10026–10036.
- (85) Fersht, A. R. Optimization of Rates of Protein Folding: The Nucleation-condensation Mechanism and Its Implications. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 10869–10873.
- (86) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. Navigating the Folding Routes. *Science* **1995**, *267*, 1619–1620.
- (87) Shakhnovich, E.; Abkevich, V.; Ptitsyn, O. Conserved Residues and the Mechanism of Protein Folding. *Nature* **1996**, *379*, 96–98.
- (88) Chen, Y.; Ding, F.; Nie, H.; Serohijos, A. W.; Sharma, S.; Wilcox, K. C.; Yin, S.; Dokholyan, N. V. Protein Folding: Then and Now. *Arch. Biochem. Biophys.* **2008**, *469*, 4–19.
- (89) Reichheld, S. E.; Yu, Z.; Davidson, A. R. The Induction of Folding Cooperativity by Ligand Binding Drives the Allosteric Response of Tetracycline Repressor. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 22263–22268.

- (90) Hatley, M. E.; Lockless, S. W.; Gibson, S. K.; Gilman, A. G.; Ranganathan, R. Allosteric Determinants in Guanine Nucleotide-binding Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 14445–14450.
- (91) Halabi, N.; Rivoire, O.; Leibler, S.; Ranganathan, R. Protein Sectors: Evolutionary Units of Three-dimensional Structure. *Cell* **2009**, *138*, 774–786.
- (92) Reynolds, K. A.; McLaughlin, R. N.; Ranganathan, R. Hot Spots for Allosteric Regulation on Protein Surfaces. *Cell* **2011**, *147*, 1564–1575.
- (93) Nussinov, R.; Ma, B.; Tsai, C. J. A Broad View of Scaffolding Suggests That Scaffolding Proteins Can Actively Control Regulation and Signaling of Multienzyme Complexes through Allostery. *Biochim. Biophys. Acta, Proteins Proteomics* **2013**, *1834*, 820–829.
- (94) Nussinov, R.; Ma, B.; Tsai, C. J. Multiple Conformational Selection and Induced Fit Events Take Place in Allosteric Propagation. *Biophys. Chem.* **2014**, *186*, 22–30.
- (95) Zhuravleva, A.; Clerico, E. M.; Giersch, L. M. An Interdomain Energetic Tug-of-war Creates the Allosterically Active State in Hsp70 Molecular Chaperones. *Cell* **2012**, *151*, 1296–1307.
- (96) Fersht, A. R.; Matouschek, A.; Serrano, L. The Folding of an Enzyme. I. Theory of Protein Engineering Analysis of Stability and Pathway of Protein Folding. *J. Mol. Biol.* **1992**, *224*, 771–782.
- (97) Nauli, S.; Kuhlman, B.; Baker, D. Computer-based Redesign of a Protein Folding Pathway. *Nat. Struct. Biol.* **2001**, *8*, 602–605.
- (98) DeGrado, W. F.; Wasserman, Z. R.; Lear, J. D. Protein Design, a Minimalist Approach. *Science* **1989**, *243*, 622–628.
- (99) Bryson, J. W.; Betz, S. F.; Lu, H. S.; Suich, D. J.; Zhou, H. X.; O’Neil, K. T.; DeGrado, W. F. Protein Design: a Hierarchic Approach. *Science* **1995**, *270*, 935–941.
- (100) Dahiyat, B. I.; Mayo, S. L. De Novo Protein Design: Fully Automated Sequence Selection. *Science* **1997**, *278*, 82–87.
- (101) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of a Novel Globular Protein Fold with Atomic-level Accuracy. *Science* **2003**, *302*, 1364–1368.
- (102) Bak, P.; Tang, C.; Wiesenfeld, K. Self-organized Criticality: An Explanation of the 1/f Noise. *Phys. Rev. Lett.* **1987**, *59*, 381–384.
- (103) Albert, R.; Jeong, H.; Barabasi, A. L. Error and Attack Tolerance of Complex Networks. *Nature* **2000**, *406*, 378–382.
- (104) Amaral, L. A.; Scala, A.; Barthelemy, M.; Stanley, H. E. Classes of Small-world Networks. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 11149–11152.
- (105) Di Paola, L.; Giuliani, A. Protein Contact Network Topology: a Natural Language for Allostery. *Curr. Opin. Struct. Biol.* **2015**, *31*, 43–48.
- (106) Di Paola, L.; De Ruvo, M.; Paci, P.; Santoni, D.; Giuliani, A. Protein Contact Networks: An Emerging Paradigm in Chemistry. *Chem. Rev.* **2013**, *113*, 1598–1613.
- (107) Rao, F.; Caflisch, A. The Protein Folding Network. *J. Mol. Biol.* **2004**, *342*, 299–306.
- (108) Serohijos, A. W.; Hegedus, T.; Riordan, J. R.; Dokholyan, N. V. Diminished Self-chaperoning Activity of the DeltaF508 Mutant of CFTR Results in Protein Misfolding. *PLoS Comput. Biol.* **2008**, *4*, e1000008.
- (109) Fersht, A. R. Characterizing Transition States in Protein Folding: An Essential Step in the Puzzle. *Curr. Opin. Struct. Biol.* **1995**, *5*, 79–84.
- (110) Vendruscolo, M.; Paci, E.; Dobson, C. M.; Karplus, M. Three Key Residues Form a Critical Contact Network in a Protein Folding Transition State. *Nature* **2001**, *409*, 641–645.
- (111) Cecconi, F.; Micheletti, C.; Carloni, P.; Maritan, A. Molecular Dynamics Studies on HIV-1 Protease Drug Resistance and Folding Pathways. *Proteins: Struct., Funct., Genet.* **2001**, *43*, 365–372.
- (112) Blacklock, K.; Verkhivker, G. M. Computational Modeling of Allosteric Regulation in the Hsp90 Chaperones: A Statistical Ensemble Analysis of Protein Structure Networks and Allosteric Communications. *PLoS Comput. Biol.* **2014**, *10*, e1003679.
- (113) Chen, J.; Dima, R. I.; Thirumalai, D. Allosteric Communication in Dihydrofolate Reductase: Signaling Network and Pathways for Closed to Occluded Transition and Back. *J. Mol. Biol.* **2007**, *374*, 250–266.
- (114) Gasper, P. M.; Fuglestad, B.; Komives, E. A.; Markwick, P. R.; McCammon, J. A. Allosteric Networks in Thrombin Distinguish Procoagulant Vs. Anticoagulant Activities. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 21216–21222.
- (115) Allain, A.; Chauvot de Beauchêne, I.; Langenfeld, F.; Guerracino, Y.; Laine, E.; Tchertanov, L. Allosteric Pathway Identification through Network Analysis: From Molecular Dynamics Simulations to Interactive 2D and 3D Graphs. *Faraday Discuss.* **2014**, *169*, 303–321.
- (116) Proctor, E. A.; Kota, P.; Aleksandrov, A. A.; He, L.; Riordan, J. R.; Dokholyan, N. V. Rational Coupled Dynamics Network Manipulation Rescues Disease-Relevant Mutant Cystic Fibrosis Transmembrane Conductance Regulator. *Chem. Sci.* **2015**, *6*, 1237–1246.
- (117) Chen, Y.; Dokholyan, N. V. Insights into Allosteric Control of Vinculin Function from Its Large Scale Conformational Dynamics. *J. Biol. Chem.* **2006**, *281*, 29148–29154.
- (118) Sharma, S.; Ding, F.; Dokholyan, N. V. Multiscale Modeling of Nucleosome Dynamics. *Biophys. J.* **2007**, *92*, 1457–1470.
- (119) Teotico, D. G.; Frazier, M. L.; Ding, F.; Dokholyan, N. V.; Temple, B. R.; Redinbo, M. R. Active Nuclear Receptors Exhibit Highly Correlated AF-2 Domain Motions. *PLoS Comput. Biol.* **2008**, *4*, e1000111.
- (120) Amadei, A.; Linssen, A. B.; Berendsen, H. J. Essential Dynamics of Proteins. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412–425.
- (121) Dokholyan, N. V.; Lee, Y.; Buldyrev, S. V.; Havlin, S.; King, P. R.; Stanley, H. E. Scaling of the Distribution of Shortest Paths in Percolation. *J. Stat. Phys.* **1998**, *93*, 603–613.
- (122) Lee, Y.; Andrade, J. S.; Buldyrev, S. V.; Dokholyan, N. V.; Havlin, S.; King, P. R.; Paul, G.; Stanley, H. E. Traveling Time and Traveling Length in Critical Percolation Clusters. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1999**, *60*, 3425–3428.
- (123) Dijkstra, E. W. A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik* **1959**, *1*, 269–271.
- (124) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. Dynamical Networks in tRNA:protein Complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6620–6625.
- (125) Bhattacharya, S.; Vaidehi, N. Differences in Allosteric Communication Pipelines in the Inactive and Active States of a GPCR. *Biophys. J.* **2014**, *107*, 422–434.
- (126) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: a Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919–11929.
- (127) Markwick, P. R.; McCammon, J. A. Studying Functional Dynamics in Bio-molecules Using Accelerated Molecular Dynamics. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20053–20065.
- (128) Krammer, A.; Lu, H.; Isralewitz, B.; Schulten, K.; Vogel, V. Forced Unfolding of the Fibronectin Type III Module Reveals a Tensile Molecular Recognition Switch. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 1351–1356.
- (129) Gao, M.; Craig, D.; Lequin, O.; Campbell, I. D.; Vogel, V.; Schulten, K. Structure and Functional Significance of Mechanically Unfolded Fibronectin Type III Intermediates. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 14784–14789.
- (130) Autenrieth, F.; Tajkhorshid, E.; Schulten, K.; Luthey-Schulten, Z. Role of Water in Transient Cytochrome c2 Docking. *J. Phys. Chem. B* **2004**, *108*, 20376–20387.
- (131) Aksimentiev, A.; Balabin, I. A.; Fillingame, R. H.; Schulten, K. Insights into the Molecular Mechanism of Rotation in the Fo Sector of ATP Synthase. *Biophys. J.* **2004**, *86*, 1332–1344.
- (132) Amaro, R. E.; Sethi, A.; Myers, R. S.; Daviesson, V. J.; Luthey-Schulten, Z. A. A Network of Conserved Interactions Regulates the Allosteric Signal in a Glutamine Amidotransferase. *Biochemistry* **2007**, *46*, 2156–2173.

- (133) Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. Discrete Molecular Dynamics Studies of the Folding of a Protein-like Model. *Folding Des.* **1998**, *3*, 577–587.
- (134) Shirvanyants, D.; Ding, F.; Tsao, D.; Ramachandran, S.; Dokholyan, N. V. Discrete Molecular Dynamics: An Efficient and Versatile Simulation Method for Fine Protein Characterization. *J. Phys. Chem. B* **2012**, *116*, 8375–8382.
- (135) Proctor, E.; Ding, F.; Dokholyan, N. Discrete Molecular Dynamics. *Comput. Mol. Sci.* **2011**, *1*, 80–92.
- (136) Alder, B. J.; Wainwright, T. E. Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.* **1959**, *31*, 459.
- (137) Zhou, Y.; Karplus, M. Folding Thermodynamics of a Model Three-helix-bundle Protein. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 14429–14432.
- (138) Zhou, Y.; Karplus, M.; Wichert, J. M.; Hall, C. K. Equilibrium Thermodynamics of Homopolymers and Clusters: Molecular Dynamics and Monte Carlo Simulations of Systems with Square-well Interactions. *J. Chem. Phys.* **1997**, *107*, 10691.
- (139) Sfriso, P.; Emperador, A.; Gelpi, J. L.; Orozco, M. Discrete Molecular Dynamics: Foundations and Biomolecular Applications. *Computational Approaches to Protein Dynamics*; CRC Press: Boca Raton, FL, 2015; pp 339–362.
- (140) Sfriso, P.; Hospital, A.; Emperador, A.; Orozco, M. Exploration of Conformational Transition Pathways from Coarse-grained Simulations. *Bioinformatics* **2013**, *29*, 1980–1986.
- (141) Ding, F.; Buldyrev, S. V.; Dokholyan, N. V. Folding Trp-cage to NMR Resolution Native Structure Using a Coarse-grained Protein Model. *Biophys. J.* **2005**, *88*, 147–155.
- (142) Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V. Ab Initio Folding of Proteins with All-atom Discrete Molecular Dynamics. *Structure* **2008**, *16*, 1010–1018.
- (143) Kota, P.; Summers, D. W.; Ren, H. Y.; Cyr, D. M.; Dokholyan, N. V. Identification of a Consensus Motif in Substrates Bound by a Type I Hsp40. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11073–11078.
- (144) Dagliyan, O.; Proctor, E. A.; D'Auria, K. M.; Ding, F.; Dokholyan, N. V. Structural and Dynamic Determinants of Protein-peptide Recognition. *Structure* **2011**, *19*, 1837–1845.
- (145) Kota, P.; García-Caballero, A.; Dang, H.; Gentzsch, M.; Stutts, M. J.; Dokholyan, N. V. Energetic and Structural Basis for Activation of the Epithelial Sodium Channel by Matriptase. *Biochemistry* **2012**, *51*, 3460–3469.
- (146) Emperador, A.; Solernou, A.; Sfriso, P.; Pons, C.; Gelpi, J. L.; Fernandez-Recio, J.; Orozco, M. Efficient Relaxation of Protein-Protein Interfaces by Discrete Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 1222–1229.
- (147) Aleksandrov, A. A.; Kota, P.; Aleksandrov, L. A.; He, L.; Jensen, T.; Cui, L.; Gentzsch, M.; Dokholyan, N. V.; Riordan, J. R. Regulatory Insertion Removal Restores Maturation, Stability and Function of DeltaF508 CFTR. *J. Mol. Biol.* **2010**, *401*, 194–210.
- (148) Serohijos, A. W.; Chen, Y.; Ding, F.; Elston, T. C.; Dokholyan, N. V. A Structural Model Reveals Energy Transduction in Dynein. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 18540–18545.
- (149) Serohijos, A. W.; Yin, S.; Ding, F.; Gauthier, J.; Gibson, D. G.; Maixner, W.; Dokholyan, N. V.; Diatchenko, L. Structural Basis for μ -Opioid Receptor Binding and Activation. *Structure* **2011**, *19*, 1683–1690.
- (150) Kota, P.; Buchner, G.; Chakraborty, H.; Dang, Y. L.; He, H.; Garcia, G. J.; Kubelka, J.; Gentzsch, M.; Stutts, M. J.; Dokholyan, N. V. The N-terminal Domain Allosterically Regulates Cleavage and Activation of the Epithelial Sodium Channel. *J. Biol. Chem.* **2014**, *289*, 23029–23042.
- (151) Tasumi, M.; Takeuchi, H.; Ataka, S.; Dwivedi, A. M.; Krimm, S. Normal Vibrations of Proteins: Glucagon. *Biopolymers* **1982**, *21*, 711–714.
- (152) Levitt, M.; Sander, C.; Stern, P. S. The Normal Modes of a Protein: Native Bovine Pancreatic Trypsin Inhibitor. *Int. J. Quantum Chem.* **1983**, *24*, 181–199.
- (153) Go, N.; Noguti, T.; Nishikawa, T. Dynamics of a Small Globular Protein in Terms of Low-frequency Vibrational Modes. *Proc. Natl. Acad. Sci. U. S. A.* **1983**, *80*, 3696–3700.
- (154) Bahar, I.; Rader, A. J. Coarse-grained Normal Mode Analysis in Structural Biology. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586–592.
- (155) Skjaerven, L.; Hollup, S. M.; Reuter, N. Normal Mode Analysis for Proteins. *J. Mol. Struct.: THEOCHEM* **2009**, *898*, 42–48.
- (156) Ma, J. Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes. *Structure* **2005**, *13*, 373–380.
- (157) Rueda, M.; Chacón, P.; Orozco, M. Thorough Validation of Protein Normal Mode Analysis: a Comparative Study with Essential Dynamics. *Structure* **2007**, *15*, 565–575.
- (158) Mouawad, L.; Perahia, D. Diagonalization in a Mixed Basis: A Method to Compute Low-frequency Normal Modes for Large Macromolecules. *Biopolymers* **1993**, *33*, 599–611.
- (159) Marques, O.; Sanejouand, Y. H. Hinge-bending Motion in Citrate Synthase Arising from Normal Mode Calculations. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 557–560.
- (160) Durand, P.; Trinquier, G.; Sanejouand, Y. A New Approach for Determining Low-Frequency Normal Modes in Macromolecules. *Biopolymers* **1994**, *34*, 759–771.
- (161) Tama, F.; Gadea, F. X.; Marques, O.; Sanejouand, Y. H. Building-block Approach for Determining Low-frequency Normal Modes of Macromolecules. *Proteins: Struct., Funct., Genet.* **2000**, *41*, 1–7.
- (162) Tirion, M. M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905–1908.
- (163) Hinsen, K. Analysis of Domain Motions by Approximate Normal Mode Calculations. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 417–429.
- (164) Bahar, I.; Atilgan, A. R.; Erman, B. Direct Evaluation of Thermal Fluctuations in Proteins Using a Single-parameter Harmonic Potential. *Folding Des.* **1997**, *2*, 173–181.
- (165) Haliloglu, T.; Bahar, I.; Erman, B. Gaussian Dynamics of Folded Proteins. *Phys. Rev. Lett.* **1997**, *79*, 3090–3093.
- (166) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophys. J.* **2001**, *80*, 505–515.
- (167) Gur, M.; Zomot, E.; Bahar, I. Global Motions Exhibited by Proteins in Micro- to Milliseconds Simulations Concur with Anisotropic Network Model Predictions. *J. Chem. Phys.* **2013**, *139*, 121912.
- (168) Vanwart, A. T.; Eargle, J.; Luthey-Schulten, Z.; Amaro, R. E. Exploring Residue Component Contributions to Dynamical Network Models of Allostery. *J. Chem. Theory Comput.* **2012**, *8*, 2949–2961.
- (169) Su, J. G.; Qi, L. S.; Li, C. H.; Zhu, Y. Y.; Du, H. J.; Hou, Y. X.; Hao, R.; Wang, J. H. Prediction of Allosteric Sites on Protein Surfaces with an Elastic-network-model-based Thermodynamic Method. *Phys. Rev. E* **2014**, *90*, 022719.
- (170) Chennubhotla, C.; Bahar, I. Markov Propagation of Allosteric Effects in Biomolecular Systems: Application to GroEL-GroES. *Mol. Syst. Biol.* **2006**, *2*, 36.
- (171) Zheng, W.; Brooks, B. R.; Thirumalai, D. Low-frequency Normal Modes That Describe Allosteric Transitions in Biological Nanomachines Are Robust to Sequence Variations. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 7664–7669.
- (172) Fenton, W. A.; Horwich, A. L. Chaperonin-mediated Protein Folding: Fate of Substrate Polypeptide. *Q. Rev. Biophys.* **2003**, *36*, 229–256.
- (173) Saibil, H. R.; Ranson, N. A. The Chaperonin Folding Machine. *Trends Biochem. Sci.* **2002**, *27*, 627–632.
- (174) Bukau, B.; Horwich, A. L. The Hsp70 and Hsp60 Chaperone Machines. *Cell* **1998**, *92*, 351–366.
- (175) Chaudhry, C.; Farr, G. W.; Todd, M. J.; Rye, H. S.; Brunger, A. T.; Adams, P. D.; Horwich, A. L.; Sigler, P. B. Role of the Gamma-phosphate of ATP in Triggering Protein Folding by GroEL-GroES: Function, Structure and Energetics. *EMBO J.* **2003**, *22*, 4877–4887.

- (176) Yifrach, O.; Horovitz, A. Nested Cooperativity in the ATPase Activity of the Oligomeric Chaperonin GroEL. *Biochemistry* **1995**, *34*, 5303–5308.
- (177) Yifrach, O.; Horovitz, A. Mapping the Transition State of the Allosteric Pathway of GroEL by Protein Engineering. *J. Am. Chem. Soc.* **1998**, *120*, 13262–13263.
- (178) Ma, J.; Karplus, M. The Allosteric Mechanism of the Chaperonin GroEL: A Dynamic Analysis. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 8502–8507.
- (179) Ditzel, L.; Löwe, J.; Stock, D.; Stetter, K. O.; Huber, H.; Huber, R.; Steinbacher, S. Crystal Structure of the Thermosome, the Archaeal Chaperonin and Homolog of CCT. *Cell* **1998**, *93*, 125–138.
- (180) Leitner, A.; Joachimiak, L. A.; Bracher, A.; Mönkemeyer, L.; Walzthoeni, T.; Chen, B.; Pechmann, S.; Holmes, S.; Cong, Y.; Ma, B.; Ludtke, S.; Chiu, W.; Hartl, F. U.; Aebersold, R.; Frydman, J. The Molecular Architecture of the Eukaryotic Chaperonin TRiC/CCT. *Structure* **2012**, *20*, 814–825.
- (181) Dekker, C.; Roe, S. M.; McCormack, E. A.; Beuron, F.; Pearl, L. H.; Willison, K. R. The Crystal Structure of Yeast CCT Reveals Intrinsic Asymmetry of Eukaryotic Cytosolic Chaperonins. *EMBO J.* **2011**, *30*, 3078–3090.
- (182) Rivenzon-Segal, D.; Wolf, S. G.; Shimon, L.; Willison, K. R.; Horovitz, A. Sequential ATP-induced Allosteric Transitions of the Cytoplasmic Chaperonin Containing TCP-1 Revealed by EM Analysis. *Nat. Struct. Mol. Biol.* **2005**, *12*, 233–237.
- (183) Koshland, D. E.; Némethy, G.; Filmer, D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* **1966**, *5*, 365–385.
- (184) Keskin, O.; Bahar, I.; Flatow, D.; Covell, D. G.; Jernigan, R. L. Molecular Mechanisms of Chaperonin GroEL-GroES Function. *Biochemistry* **2002**, *41*, 491–501.
- (185) Zheng, W.; Brooks, B. R.; Thirumalai, D. Allosteric Transitions in the Chaperonin GroEL Are Captured by a Dominant Normal Mode That Is Most Robust to Sequence Variations. *Biophys. J.* **2007**, *93*, 2289–2299.
- (186) Chennubhotla, C.; Yang, Z.; Bahar, I. Coupling Between Global Dynamics and Signal Transduction Pathways: a Mechanism of Allostery for Chaperonin GroEL. *Mol. BioSyst.* **2008**, *4*, 287–292.
- (187) Tehver, R.; Chen, J.; Thirumalai, D. Allostery Wiring Diagrams in the Transitions That Drive the GroEL Reaction Cycle. *J. Mol. Biol.* **2009**, *387*, 390–406.
- (188) Jayasinghe, M.; Shrestha, P.; Wu, X.; Tehver, R.; Stan, G. Weak Intra-ring Allosteric Communications of the Archaeal Chaperonin Thermosome Revealed by Normal Mode Analysis. *Biophys. J.* **2012**, *103*, 1285–1295.
- (189) Doyle, D. A.; Lee, A.; Lewis, J.; Kim, E.; Sheng, M.; MacKinnon, R. Crystal Structures of a Complexed and Peptide-free Membrane Protein-binding Domain: Molecular Basis of Peptide Recognition by PDZ. *Cell* **1996**, *85*, 1067–1076.
- (190) Süel, G. M.; Lockless, S. W.; Wall, M. A.; Ranganathan, R. Evolutionarily Conserved Networks of Residues Mediate Allosteric Communication in Proteins. *Nat. Struct. Biol.* **2003**, *10*, 59–69.
- (191) Dima, R. I.; Thirumalai, D. Determination of Network of Residues That Regulate Allostery in Protein Families Using Sequence Analysis. *Protein Sci.* **2006**, *15*, 258–268.
- (192) Fodor, A. A.; Aldrich, R. W. Influence of Conservation on Calculations of Amino Acid Covariance in Multiple Sequence Alignments. *Proteins: Struct., Funct., Genet.* **2004**, *56*, 211–221.
- (193) Kass, I.; Horovitz, A. Mapping Pathways of Allosteric Communication in GroEL by Analysis of Correlated Mutations. *Proteins: Struct., Funct., Genet.* **2002**, *48*, 611–617.
- (194) Atchley, W. R.; Wollenberg, K. R.; Fitch, W. M.; Terhalle, W.; Dress, A. W. Correlations Among Amino Acid Sites in bHLH Protein Domains: An Information Theoretic Analysis. *Mol. Biol. Evol.* **2000**, *17*, 164–178.
- (195) Göbel, U.; Sander, C.; Schneider, R.; Valencia, A. Correlated Mutations and Residue Contacts in Proteins. *Proteins: Struct., Funct., Genet.* **1994**, *18*, 309–317.
- (196) Olmea, O.; Rost, B.; Valencia, A. Effective Use of Sequence Correlation and Conservation in Fold Recognition. *J. Mol. Biol.* **1999**, *293*, 1221–1239.
- (197) Fodor, A. A.; Aldrich, R. W. On Evolutionary Conservation of Thermodynamic Coupling in Proteins. *J. Biol. Chem.* **2004**, *279*, 19046–19050.
- (198) Marks, D. S.; Hopf, T. A.; Sander, C. Protein Structure Prediction from Sequence Variation. *Nat. Biotechnol.* **2012**, *30*, 1072–1080.
- (199) Marks, D. S.; Colwell, L. J.; Sheridan, R.; Hopf, T. A.; Pagnani, A.; Zecchina, R.; Sander, C. Protein 3D Structure Computed from Evolutionary Sequence Variation. *PLoS One* **2011**, *6*, e28766.
- (200) Sulkowska, J. I.; Morcos, F.; Weigt, M.; Hwa, T.; Onuchic, J. N. Genomics-aided Structure Prediction. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 10340–10345.
- (201) Nugent, T.; Jones, D. T. Accurate de Novo Structure Prediction of Large Transmembrane Protein Domains Using Fragment-assembly and Correlated Mutation Analysis. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E1540–E1547.
- (202) Wollenberg, K. R.; Atchley, W. R. Separation of Phylogenetic and Functional Associations in Biological Sequences by Using the Parametric Bootstrap. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 3288–3291.
- (203) Noivirt, O.; Eisenstein, M.; Horovitz, A. Detection and Reduction of Evolutionary Noise in Correlated Mutation Analysis. *Protein Eng., Des. Sel.* **2005**, *18*, 247–253.
- (204) Pei, J.; Dokholyan, N. V.; Shakhnovich, E. I.; Grishin, N. V. Using Protein Design for Homology Detection and Active Site Searches. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 11361–11366.
- (205) Hudson, J. W.; Golding, G. B.; Crerar, M. M. Evolution of Allosteric Control in Glycogen Phosphorylase. *J. Mol. Biol.* **1993**, *234*, 700–721.
- (206) Jensen, A. A.; Spalding, T. A. Allosteric Modulation of G-protein Coupled Receptors. *Eur. J. Pharm. Sci.* **2004**, *21*, 407–420.
- (207) May, L. T.; Leach, K.; Sexton, P. M.; Christopoulos, A. Allosteric Modulation of G Protein-coupled Receptors. *Annu. Rev. Pharmacol. Toxicol.* **2007**, *47*, 1–51.
- (208) Convertino, M.; Samoshkin, A.; Viet, C. T.; Gauthier, J.; Li Fraire, S. P.; Sharif-Naeini, R.; Schmidt, B. L.; Maixner, W.; Diatchenko, L.; Dokholyan, N. V. Differential Regulation of 6- and 7-Transmembrane Helix Variants of μ -Opioid Receptor in Response to Morphine Stimulation. *PLoS One* **2015**, *10*, e0142826.
- (209) Samoshkin, A.; Convertino, M.; Viet, C. T.; Wieskopf, J. S.; Kambur, O.; Marcovitz, J.; Patel, P.; Stone, L. S.; Kalso, E.; Mogil, J. S.; Schmidt, B. L.; Maixner, W.; Dokholyan, N. V.; Diatchenko, L. Structural and Functional Interactions Between Six-transmembrane M-opioid Receptors and B2-adrenoreceptors Modulate Opioid Signaling. *Sci. Rep.* **2015**, *5*, 18198.
- (210) Livesay, D. R.; Kreth, K. E.; Fodor, A. A. A Critical Evaluation of Correlated Mutation Algorithms and Coevolution Within Allosteric Mechanisms. *Methods Mol. Biol.* **2012**, *796*, 385–398.
- (211) Royer, W. E.; Knapp, J. E.; Strand, K.; Heaslet, H. A. Cooperative Hemoglobins: Conserved Fold, Diverse Quaternary Assemblies and Allosteric Mechanisms. *Trends Biochem. Sci.* **2001**, *26*, 297–304.
- (212) Royer, W. E.; Zhu, H.; Gorr, T. A.; Flores, J. F.; Knapp, J. E. Allosteric Hemoglobin Assembly: Diversity and Similarity. *J. Biol. Chem.* **2005**, *280*, 27477–27480.
- (213) Kuriyan, J.; Eisenberg, D. The Origin of Protein Interactions and Allostery in Colocalization. *Nature* **2007**, *450*, 983–990.
- (214) Del Sol, A.; Tsai, C. J.; Ma, B.; Nussinov, R. The Origin of Allosteric Functional Modulation: Multiple Pre-existing Pathways. *Structure* **2009**, *17*, 1042–1050.
- (215) Formaneck, M. S.; Ma, L.; Cui, Q. Reconciling the “Old” and “New” Views of Protein Allostery: a Molecular Simulation Study of Chemotaxis Y Protein (CheY). *Proteins: Struct., Funct., Genet.* **2006**, *63*, 846–867.
- (216) Liu, Z.; Chen, J.; Thirumalai, D. On the Accuracy of Inferring Energetic Coupling Between Distant Sites in Protein Families from

- Evolutionary Imprints: Illustrations Using Lattice Model. *Proteins: Struct., Funct., Genet.* **2009**, *77*, 823–831.
- (217) Tang, S.; Liao, J. C.; Dunn, A. R.; Altman, R. B.; Spudich, J. A.; Schmidt, J. P. Predicting Allosteric Communication in Myosin via a Pathway of Conserved Residues. *J. Mol. Biol.* **2007**, *373*, 1361–1373.
- (218) Mittermaier, A.; Kay, L. E. New Tools Provide New Insights in NMR Studies of Protein Dynamics. *Science* **2006**, *312*, 224–228.
- (219) Smock, R. G.; Giersch, L. M. Sending Signals Dynamically. *Science* **2009**, *324*, 198–203.
- (220) Nussinov, R.; Tsai, C. J. Allostery in Disease and in Drug Discovery. *Cell* **2013**, *153*, 293–305.
- (221) Das, R.; Esposito, V.; Abu-Abed, M.; Anand, G. S.; Taylor, S. S.; Melacini, G. cAMP Activation of PKA Defines an Ancient Signaling Mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 93–98.
- (222) Manley, G.; Loria, J. P. NMR Insights into Protein Allostery. *Arch. Biochem. Biophys.* **2012**, *519*, 223–231.
- (223) Masterson, L. R.; Mascioni, A.; Traaseth, N. J.; Taylor, S. S.; Veglia, G. Allosteric Cooperativity in Protein Kinase A. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 506–511.
- (224) Tzeng, S. R.; Kalodimos, C. G. Allosteric Inhibition through Suppression of Transient Conformational States. *Nat. Chem. Biol.* **2013**, *9*, 462–465.
- (225) Shi, L.; Kay, L. E. Tracing an Allosteric Pathway Regulating the Activity of the HsLV Protease. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 2140–2145.
- (226) Clarkson, M. W.; Gilmore, S. A.; Edgell, M. H.; Lee, A. L. Dynamic Coupling and Allosteric Behavior in a Nonallosteric Protein. *Biochemistry* **2006**, *45*, 7693–7699.
- (227) Fuentes, E. J.; Gilmore, S. A.; Mauldin, R. V.; Lee, A. L. Evaluation of Energetic and Dynamic Coupling Networks in a PDZ Domain Protein. *J. Mol. Biol.* **2006**, *364*, 337–351.
- (228) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. Protein Structure Determination from NMR Chemical Shifts. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 9615–9620.
- (229) Camilloni, C.; Robustelli, P.; De Simone, A.; Cavalli, A.; Vendruscolo, M. Characterization of the Conformational Equilibrium Between the Two Major Substates of RNase A Using NMR Chemical Shifts. *J. Am. Chem. Soc.* **2012**, *134*, 3968–3971.
- (230) Robustelli, P.; Kohlhoff, K.; Cavalli, A.; Vendruscolo, M. Using NMR Chemical Shifts as Structural Restraints in Molecular Dynamics Simulations of Proteins. *Structure* **2010**, *18*, 923–933.
- (231) Camilloni, C.; Cavalli, A.; Vendruscolo, M. Assessment of the Use of NMR Chemical Shifts as Replica-averaged Structural Restraints in Molecular Dynamics Simulations to Characterize the Dynamics of Proteins. *J. Phys. Chem. B* **2013**, *117*, 1838–1843.
- (232) De Simone, A.; Montalvo, R. W.; Dobson, C. M.; Vendruscolo, M. Characterization of the Interdomain Motions in Hen Lysozyme Using Residual Dipolar Couplings as Replica-averaged Structural Restraints in Molecular Dynamics Simulations. *Biochemistry* **2013**, *52*, 6480–6486.
- (233) Best, R. B.; Vendruscolo, M. Structural Interpretation of Hydrogen Exchange Protection Factors in Proteins: Characterization of the Native State Fluctuations of CI2. *Structure* **2006**, *14*, 97–106.
- (234) Dixon, R. D.; Chen, Y.; Ding, F.; Khare, S. D.; Prutzman, K. C.; Schaller, M. D.; Campbell, S. L.; Dokholyan, N. V. New Insights into FAK Signaling and Localization Based on Detection of a FAT Domain Folding Intermediate. *Structure* **2004**, *12*, 2161–2171.
- (235) Chen, Y.; Campbell, S. L.; Dokholyan, N. V. Deciphering Protein Dynamics from NMR Data Using Explicit Structure Sampling and Selection. *Biophys. J.* **2007**, *93*, 2300–2306.
- (236) Grosman, C.; Zhou, M.; Auerbach, A. Mapping the Conformational Wave of Acetylcholine Receptor Channel Gating. *Nature* **2000**, *403*, 773–776.
- (237) Mitra, A.; Tascione, R.; Auerbach, A.; Licht, S. Plasticity of Acetylcholine Receptor Gating Motions via Rate-energy Relationships. *Biophys. J.* **2005**, *89*, 3071–3078.
- (238) Leffler, J. E. Parameters for the Description of Transition States. *Science* **1953**, *117*, 340–341.
- (239) Oliveberg, M.; Tan, Y. J.; Silow, M.; Fersht, A. R. The Changing Nature of the Protein Folding Transition State: Implications for the Shape of the Free-energy Profile for Folding. *J. Mol. Biol.* **1998**, *277*, 933–943.
- (240) Jackson, S. E. How Do Small Single-domain Proteins Fold? *Folding Des.* **1998**, *3*, R81–R91.
- (241) Salvatella, X.; Dobson, C. M.; Fersht, A. R.; Vendruscolo, M. Determination of the Folding Transition States of Barnase by Using PhiI-value-restrained Simulations Validated by Double Mutant PhiIJ-values. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 12389–12394.
- (242) Hubner, I. A.; Edmonds, K. A.; Shakhnovich, E. I. Nucleation and the Transition State of the SH3 Domain. *J. Mol. Biol.* **2005**, *349*, 424–434.
- (243) Levy, Y.; Cho, S. S.; Onuchic, J. N.; Wolynes, P. G. A Survey of Flexible Protein Binding Mechanisms and Their Transition States Using Native Topology Based Energy Landscapes. *J. Mol. Biol.* **2005**, *346*, 1121–1145.
- (244) Matouschek, A.; Kellis, J. T.; Serrano, L.; Fersht, A. R. Mapping the Transition State and Pathway of Protein Folding by Protein Engineering. *Nature* **1989**, *340*, 122–126.
- (245) Matouschek, A.; Kellis, J. T.; Serrano, L.; Bycroft, M.; Fersht, A. R. Transient Folding Intermediates Characterized by Protein Engineering. *Nature* **1990**, *346*, 440–445.
- (246) Horovitz, A.; Amir, A.; Danziger, O.; Kafri, G. Phi Value Analysis of Heterogeneity in Pathways of Allosteric Transitions: Evidence for Parallel Pathways of ATP-induced Conformational Changes in a GroEL Ring. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 14095–14097.
- (247) Serohijos, A. W.; Hegedus, T.; Aleksandrov, A. A.; He, L.; Cui, L.; Dokholyan, N. V.; Riordan, J. R. Phenylalanine-508 Mediates a Cytoplasmic-membrane Domain Contact in the CFTR 3D Structure Crucial to Assembly and Channel Function. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 3256–3261.
- (248) Hultqvist, G.; Haq, S. R.; Punekar, A. S.; Chi, C. N.; Engström, Å.; Bach, A.; Stroemgaard, K.; Selmer, M.; Gianni, S.; Jemth, P. Energetic Pathway Sampling in a Protein Interaction Domain. *Structure* **2013**, *21*, 1193–1202.
- (249) Schreiber, G.; Fersht, A. R. Energetics of Protein-protein Interactions: Analysis of the Barnase-barstar Interface by Single Mutations and Double Mutant Cycles. *J. Mol. Biol.* **1995**, *248*, 478–486.
- (250) Kohlstaedt, L. A.; Wang, J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. Crystal Structure at 3.5 Å Resolution of HIV-1 Reverse Transcriptase Complexed with an Inhibitor. *Science* **1992**, *256*, 1783–1790.
- (251) Oikonomakos, N. G.; Skamnaki, V. T.; Tsitsanou, K. E.; Gavalas, N. G.; Johnson, L. N. A New Allosteric Site in Glycogen Phosphorylase b as a Target for Drug Interactions. *Structure* **2000**, *8*, 575–584.
- (252) Rath, V. L.; Ammirati, M.; Danley, D. E.; Ekstrom, J. L.; Gibbs, E. M.; Hynes, T. R.; Mathiowitz, A. M.; McPherson, R. K.; Olson, T. V.; Treadway, J. L.; Hoover, D. J. Human Liver Glycogen Phosphorylase Inhibitors Bind at a New Allosteric Site. *Chem. Biol.* **2000**, *7*, 677–682.
- (253) Hardy, J. A.; Wells, J. A. Searching for New Allosteric Sites in Enzymes. *Curr. Opin. Struct. Biol.* **2004**, *14*, 706–715.
- (254) Hardy, J. A.; Lam, J.; Nguyen, J. T.; O'Brien, T.; Wells, J. A. Discovery of an Allosteric Site in the Caspases. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 12461–12466.
- (255) Radley, T. L.; Markowska, A. I.; Bettinger, B. T.; Ha, J. H.; Loh, S. N. Allosteric Switching by Mutually Exclusive Folding of Protein Domains. *J. Mol. Biol.* **2003**, *332*, 529–536.
- (256) Ha, J. H.; Butler, J. S.; Mitrea, D. M.; Loh, S. N. Modular Enzyme Design: Regulation by Mutually Exclusive Protein Folding. *J. Mol. Biol.* **2006**, *357*, 1058–1062.
- (257) Guntas, G.; Mitchell, S. F.; Ostermeier, M. A Molecular Switch Created by in Vitro Recombination of Nonhomologous Genes. *Chem. Biol.* **2004**, *11*, 1483–1487.

- (258) Guntas, G.; Ostermeier, M. Creation of an Allosteric Enzyme by Domain Insertion. *J. Mol. Biol.* **2004**, *336*, 263–273.
- (259) Guntas, G.; Mansell, T. J.; Kim, J. R.; Ostermeier, M. Directed Evolution of Protein Switches and Their Application to the Creation of Ligand-binding Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 11224–11229.
- (260) Choi, J. H.; Laurent, A. H.; Hilser, V. J.; Ostermeier, M. Design of Protein Switches Based on an Ensemble Model of Allostery. *Nat. Commun.* **2015**, *6*, 6968.
- (261) Tucker, C. L.; Fields, S. A Yeast Sensor of Ligand Binding. *Nat. Biotechnol.* **2001**, *19*, 1042–1046.
- (262) Karginov, A. V.; Zou, Y.; Shirvanyants, D.; Kota, P.; Dokholyan, N. V.; Young, D. D.; Hahn, K. M.; Deiters, A. Light Regulation of Protein Dimerization and Kinase Activity in Living Cells Using Photocaged Rapamycin and Engineered FKBP. *J. Am. Chem. Soc.* **2011**, *133*, 420–423.
- (263) Liang, J.; Choi, J.; Clardy, J. Refined Structure of the FKBP12-rapamycin-FRB Ternary Complex at 2.2 Å Resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, *55*, 736–744.
- (264) Karginov, A. V.; Ding, F.; Kota, P.; Dokholyan, N. V.; Hahn, K. M. Engineered Allosteric Activation of Kinases in Living Cells. *Nat. Biotechnol.* **2010**, *28*, 743–747.
- (265) Chu, P. H.; Tsygankov, D.; Berginski, M. E.; Dagliyan, O.; Gomez, S. M.; Elston, T. C.; Karginov, A. V.; Hahn, K. M. Engineered Kinase Activation Reveals Unique Morphodynamic Phenotypes and Associated Trafficking for Src Family Isoforms. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 12420–12425.
- (266) Brown, K. A.; Zou, Y.; Shirvanyants, D.; Zhang, J.; Samanta, S.; Mantravadi, P. K.; Dokholyan, N. V.; Deiters, A. Light-cleavable Rapamycin Dimer as an Optical Trigger for Protein Dimerization. *Chem. Commun.* **2015**, *51*, 5702–5705.
- (267) Dagliyan, O.; Shirvanyants, D.; Karginov, A. V.; Ding, F.; Fee, L.; Chandrasekaran, S. N.; Freisinger, C. M.; Smolen, G. A.; Huttenlocher, A.; Hahn, K. M.; Dokholyan, N. V. Rational Design of a Ligand-controlled Protein Conformational Switch. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6800–6804.
- (268) Ding, F.; Dokholyan, N. V. Emergence of Protein Fold Families through Rational Design. *PLoS Comput. Biol.* **2006**, *2*, e85.
- (269) Yin, S.; Ding, F.; Dokholyan, N. V. Modeling Backbone Flexibility Improves Protein Stability Estimation. *Structure* **2007**, *15*, 1567–1576.
- (270) Yin, S.; Ding, F.; Dokholyan, N. Eris: An Automated Estimator of Protein Stability. *Nat. Methods* **2007**, *4*, 466–467.
- (271) Holm, L. Unification of Protein Families. *Curr. Opin. Struct. Biol.* **1998**, *8*, 372–379.
- (272) Holmes, K. C.; Sander, C.; Valencia, A. A New ATP-binding Fold in Actin, Hexokinase and Hsc70. *Trends Cell Biol.* **1993**, *3*, 53–59.
- (273) Orengo, C. A.; Michie, A. D.; Jones, S.; Jones, D. T.; Swindells, M. B.; Thornton, J. M. CATH—a Hierarchic Classification of Protein Domain Structures. *Structure* **1997**, *5*, 1093–1108.
- (274) Dodge, C.; Schneider, R.; Sander, C. The HSSP Database of Protein Structure-sequence Alignments and Family Profiles. *Nucleic Acids Res.* **1998**, *26*, 313–315.
- (275) Sánchez, R.; Pieper, U.; Melo, F.; Eswar, N.; Martí-Renom, M. A.; Madhusudhan, M. S.; Mirković, N.; Sali, A. Protein Structure Modeling for Structural Genomics. *Nat. Struct. Biol.* **2000**, *7*, 986–990.
- (276) Pearl, F. M.; Lee, D.; Bray, J. E.; Sillitoe, I.; Todd, A. E.; Harrison, A. P.; Thornton, J. M.; Orengo, C. A. Assigning Genomic Sequences to CATH. *Nucleic Acids Res.* **2000**, *28*, 277–282.
- (277) Dalal, S.; Regan, L. Understanding the Sequence Determinants of Conformational Switching Using Protein Design. *Protein Sci.* **2000**, *9*, 1651–1659.
- (278) Cordes, M. H.; Burton, R. E.; Walsh, N. P.; McKnight, C. J.; Sauer, R. T. An Evolutionary Bridge to a New Protein Fold. *Nat. Struct. Biol.* **2000**, *7*, 1129–1132.
- (279) Cordes, M. H.; Walsh, N. P.; McKnight, C. J.; Sauer, R. T. Evolution of a Protein Fold in Vitro. *Science* **1999**, *284*, 325–327.
- (280) Dokholyan, N. V. The Architecture of the Protein Domain Universe. *Gene* **2005**, *347*, 199–206.
- (281) LeFevre, K. R.; Cordes, M. H. Retroevolution of Lambda Cro Toward a Stable Monomer. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 2345–2350.
- (282) Dokholyan, N. V. Protein Designability and Engineering. In *Structural Bioinformatics*; Bourne, P. E., Gu, J., Eds.; Wiley & Sons: Hoboken, NJ, 2008; pp 961–982.
- (283) Dado, G. P.; Gellman, S. H. Redox Control of Secondary Structure in a Designed Peptide. *J. Am. Chem. Soc.* **1993**, *115*, 12609–12610.
- (284) Altman, M.; Lee, P.; Rich, A.; Zhang, S. Conformational Behavior of Ionic Self-complementary Peptides. *Protein Sci.* **2000**, *9*, 1095–1105.
- (285) Cerpa, R.; Cohen, F. E.; Kuntz, I. D. Conformational Switching in Designed Peptides: The Helix/sheet Transition. *Folding Des.* **1996**, *1*, 91–101.
- (286) Fezoui, Y.; Hartley, D. M.; Walsh, D. M.; Selkoe, D. J.; Osterhout, J. J.; Teplow, D. B. A de Novo Designed Helix-turn-helix Peptide Forms Nontoxic Amyloid Fibrils. *Nat. Struct. Biol.* **2000**, *7*, 1095–1099.
- (287) Grishin, N. V. Fold Change in Evolution of Protein Structures. *J. Struct. Biol.* **2001**, *134*, 167–185.
- (288) Lee, H. M.; Chmielewski, J. Liposomal Cargo Unloading Induced by pH-sensitive Peptides. *J. Pept. Res.* **2005**, *65*, 355–363.
- (289) Li, W.; Nicol, F.; Szoka, F. C. GALA: a Designed Synthetic pH-responsive Amphiphatic Peptide with Applications in Drug and Gene Delivery. *Adv. Drug Delivery Rev.* **2004**, *56*, 967–985.
- (290) Mutter, M.; Gassmann, R.; Buttkus, U.; Altmann, K.-H. Switch Peptides: pH-Induced α -Helix to β -Sheet Transitions of Bis-amphiphilic Oligopeptides. *Angew. Chem., Int. Ed. Engl.* **1991**, *30*, 1514–1516.
- (291) Ciani, B.; Hutchinson, E. G.; Sessions, R. B.; Woolfson, D. N. A Designed System for Assessing How Sequence Affects Alpha to Beta Conformational Transitions in Proteins. *J. Biol. Chem.* **2002**, *277*, 10150–10155.
- (292) Kammerer, R. A.; Kostrewa, D.; Zurdo, J.; Detken, A.; García-Echeverría, C.; Green, J. D.; Müller, S. A.; Meier, B. H.; Winkler, F. K.; Dobson, C. M.; Steinmetz, M. O. Exploring Amyloid Formation by a de Novo Design. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 4435–4440.
- (293) Pagel, K.; Vagt, T.; Kohajda, T.; Koksch, B. From Alpha-helix to Beta-sheet—a Reversible Metal Ion Induced Peptide Secondary Structure Switch. *Org. Biomol. Chem.* **2005**, *3*, 2500–2502.
- (294) Cerasoli, E.; Sharpe, B. K.; Woolfson, D. N. ZiCo: a Peptide Designed to Switch Folded State Upon Binding Zinc. *J. Am. Chem. Soc.* **2005**, *127*, 15008–15009.
- (295) Ambroggio, X. I.; Kuhlman, B. Computational Design of a Single Amino Acid Sequence That Can Switch Between Two Distinct Protein Folds. *J. Am. Chem. Soc.* **2006**, *128*, 1154–1161.
- (296) Bullough, P. A.; Hughson, F. M.; Skehel, J. J.; Wiley, D. C. Structure of Influenza Haemagglutinin at the pH of Membrane Fusion. *Nature* **1994**, *371*, 37–43.
- (297) Elrod-Erickson, M.; Rould, M. A.; Nekludova, L.; Pabo, C. O. Zif268 protein-DNA Complex Refined at 1.6 Å: a Model System for Understanding Zinc finger-DNA Interactions. *Structure* **1996**, *4*, 1171–1180.
- (298) Kuhlman, B.; Baker, D. Native Protein Sequences Are Close to Optimal for Their Structures. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 10383–10388.
- (299) Ambroggio, X. I.; Kuhlman, B. Design of Protein Conformational Switches. *Curr. Opin. Struct. Biol.* **2006**, *16*, 525–530.
- (300) Wright, C. M.; Heins, R. A.; Ostermeier, M. As Easy as Flipping a Switch? *Curr. Opin. Chem. Biol.* **2007**, *11*, 342–346.
- (301) Cross, P. J.; Allison, T. M.; Dobson, R. C.; Jameson, G. B.; Parker, E. J. Engineering Allosteric Control to an Unregulated Enzyme by Transfer of a Regulatory Domain. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 2111–2116.
- (302) Curien, G.; Biou, V.; Mas-Droux, C.; Robert-Genthon, M.; Ferrer, J. L.; Dumas, R. Amino Acid Biosynthesis: New Architectures in Allosteric Enzymes. *Plant Physiol. Biochem.* **2008**, *46*, 325–339.

- (303) Raman, S.; Taylor, N.; Genuth, N.; Fields, S.; Church, G. M. Engineering Allostery. *Trends Genet.* **2014**, *30*, 521–528.
- (304) Dietrich, J. A.; Shis, D. L.; Alikhani, A.; Keasling, J. D. Transcription Factor-based Screens and Synthetic Selections for Microbial Small-molecule Biosynthesis. *ACS Synth. Biol.* **2013**, *2*, 47–58.
- (305) Fowler, D. M.; Fields, S. Deep Mutational Scanning: a New Style of Protein Science. *Nat. Methods* **2014**, *11*, 801–807.
- (306) Strickland, D.; Moffat, K.; Sosnick, T. R. Light-activated DNA Binding in a Designed Allosteric Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 10709–10714.
- (307) Salomon, M.; Eisenreich, W.; Dürr, H.; Schleicher, E.; Knieb, E.; Massey, V.; Rüdiger, W.; Müller, F.; Bacher, A.; Richter, G. An Optomechanical Transducer in the Blue Light Receptor Phototropin from *Avena Sativa*. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 12357–12361.
- (308) Cosentino, C.; Alberio, L.; Gazzarrini, S.; Aquila, M.; Romano, E.; Cermenati, S.; Zuccolini, P.; Petersen, J.; Beltrame, M.; Van Etten, J. L.; Christie, J. M.; Thiel, G.; Moroni, A. Optogenetics. Engineering of a Light-gated Potassium Channel. *Science* **2015**, *348*, 707–710.
- (309) Lee, J.; Natarajan, M.; Nashine, V. C.; Socolich, M.; Vo, T.; Russ, W. P.; Benkovic, S. J.; Ranganathan, R. Surface Sites for Engineering Allosteric Control in Proteins. *Science* **2008**, *322*, 438–442.
- (310) Riordan, J. R.; Rommens, J. M.; Kerem, B.; Alon, N.; Rozmahel, R.; Grzelczak, Z.; Zielenski, J.; Lok, S.; Plavsic, N.; Chou, J. L.; et al. Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA. *Science* **1989**, *245*, 1066–1073.
- (311) Quinton, P. M. Cystic Fibrosis: Lessons from the Sweat Gland. *Physiology* **2007**, *22*, 212–225.
- (312) Davies, J. C.; Alton, E. W.; Bush, A. Cystic Fibrosis. *Br. Med. J.* **2007**, *335*, 1255–1259.
- (313) Riordan, J. R. CFTR Function and Prospects for Therapy. *Annu. Rev. Biochem.* **2008**, *77*, 701–726.
- (314) Chong, P. A.; Kota, P.; Dokholyan, N. V.; Forman-Kay, J. D. Dynamics Intrinsic to Cystic Fibrosis Transmembrane Conductance Regulator Function and Stability. *Cold Spring Harbor Perspect. Med.* **2013**, *3*, a009522.
- (315) He, L.; Aleksandrov, A. A.; An, J.; Cui, L.; Yang, Z.; Brouillette, C. G.; Riordan, J. R. Restoration of NBD1 Thermal Stability Is Necessary and Sufficient to Correct ΔF508 CFTR Folding and Assembly. *J. Mol. Biol.* **2015**, *427*, 106–120.
- (316) Ding, F.; Dokholyan, N. V. Simple but Predictive Protein Models. *Trends Biotechnol.* **2005**, *23*, 450–455.
- (317) Aleksandrov, A. A.; Kota, P.; Cui, L.; Jensen, T.; Alekseev, A. E.; Reyes, S.; He, L.; Gentzsch, M.; Aleksandrov, L. A.; Dokholyan, N. V.; Riordan, J. R. Allosteric Modulation Balances Thermodynamic Stability and Restores Function of ΔF508 CFTR. *J. Mol. Biol.* **2012**, *419*, 41–60.
- (318) Nilsen, T. W. RNA 1997–2007: a Remarkable Decade of Discovery. *Mol. Cell* **2007**, *28*, 715–720.
- (319) Cruz, J. A.; Westhof, E. The Dynamic Landscapes of RNA Architecture. *Cell* **2009**, *136*, 604–609.
- (320) Sharp, P. A. The Centrality of RNA. *Cell* **2009**, *136*, 577–580.
- (321) Wan, Y.; Kertesz, M.; Spitale, R. C.; Segal, E.; Chang, H. Y. Understanding the Transcriptome through RNA Structure. *Nat. Rev. Genet.* **2011**, *12*, 641–655.
- (322) Cech, T. R.; Steitz, J. A. The Noncoding RNA Revolution—trashing Old Rules to Forge New Ones. *Cell* **2014**, *157*, 77–94.
- (323) Sharma, S.; Ding, F.; Nie, H.; Watson, D.; Unnithan, A.; Lopp, J.; Pozefsky, D.; Dokholyan, N. V. iFold: a Platform for Interactive Folding Simulations of Proteins. *Bioinformatics* **2006**, *22*, 2693–2694.
- (324) Gherghe, C. M.; Leonard, C. W.; Ding, F.; Dokholyan, N. V.; Weeks, K. M. Native-like RNA Tertiary Structures Using a Sequence-encoded Cleavage Agent and Refinement by Discrete Molecular Dynamics. *J. Am. Chem. Soc.* **2009**, *131*, 2541–2546.
- (325) Das, R.; Karanicolas, J.; Baker, D. Atomic Accuracy in Predicting and Designing Noncanonical RNA Structure. *Nat. Methods* **2010**, *7*, 291–294.
- (326) Kladwang, W.; VanLang, C. C.; Cordero, P.; Das, R. A Two-dimensional Mutate-and-map Strategy for Non-coding RNA Structure. *Nat. Chem.* **2011**, *3*, 954–962.
- (327) Ding, F.; Lavender, C. A.; Weeks, K. M.; Dokholyan, N. V. Three-dimensional RNA Structure Refinement by Hydroxyl Radical Probing. *Nat. Methods* **2012**, *9*, 603–608.
- (328) Zhao, Y.; Huang, Y.; Gong, Z.; Wang, Y.; Man, J.; Xiao, Y. Automated and Fast Building of Three-dimensional RNA Structures. *Sci. Rep.* **2012**, *2*, 734.
- (329) Sim, A. Y.; Minary, P.; Levitt, M. Modeling Nucleic Acids. *Curr. Opin. Struct. Biol.* **2012**, *22*, 273–278.
- (330) Karplus, M.; Weaver, D. L. Protein-folding Dynamics. *Nature* **1976**, *260*, 404–406.
- (331) Serohijos, A. W. R.; Yin, S. Y.; Ding, F.; Gauthier, J.; Gibson, D. G.; Maixner, W.; Dokholyan, N. V.; Diatchenko, L. Structural Basis for mu-Opioid Receptor Binding and Activation. *Structure* **2011**, *19*, 1683–1690.
- (332) Miao, Y.; Nichols, S. E.; Gasper, P. M.; Metzger, V. T.; McCammon, J. A. Activation and Dynamic Network of the M2Muscarinic Receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 10982–10987.
- (333) Dror, R. O.; Arlow, D. H.; Maragakis, P.; Mildorf, T. J.; Pan, A. C.; Xu, H.; Borhani, D. W.; Shaw, D. E. Activation Mechanism of the B2-adrenergic Receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 18684–18689.