



# Biophysical and computational methods to analyze amino acid interaction networks in proteins

Kathleen F. O'Rourke<sup>1</sup>, Scott D. Gorman<sup>1</sup>, David D. Boehr<sup>\*</sup>

Department of Chemistry, The Pennsylvania State University, University Park, PA 16802, USA

## ARTICLE INFO

### Article history:

Received 30 March 2016

Received in revised form 4 June 2016

Accepted 13 June 2016

Available online 22 June 2016

### Keywords:

Amino acid interaction network

Allostery

Graph theory

CONTACT

Molecular dynamics

Elastic network

NMR

Coevolution

Statistical coupling analysis

## ABSTRACT

Globular proteins are held together by interacting networks of amino acid residues. A number of different structural and computational methods have been developed to interrogate these amino acid networks. In this review, we describe some of these methods, including analyses of X-ray crystallographic data and structures, computer simulations, NMR data, and covariation among protein sequences, and indicate the critical insights that such methods provide into protein function. This information can be leveraged towards the design of new allosteric drugs, and the engineering of new protein function and protein regulation strategies.

© 2016 O'Rourke et al. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

It has long been understood that interactions at the local level (e.g. H-bonding, steric interactions) dictate the formation of protein structural elements, such as  $\alpha$ -helices and  $\beta$ -sheets, and that local interactions also dictate the packing of these various structural elements to form three-dimensional protein structure (e.g. ref. [1,2]). There is also now a better appreciation for the local interactions that are important for loop structure and dynamics (e.g. ref. [3]). With these energetic considerations in mind, globular proteins can be viewed as being held together by a series of local interactions through networks of interacting amino acid residues. These amino acid networks (Fig. 1) have also been termed 'residue interaction networks' [4], 'protein structure networks' [5], 'contact networks' [6], 'pathways' [7], 'circuits' [8], 'wiring diagrams' [9], 'protein sectors' [10] and so on. Intrinsic to this viewpoint is the idea that some interactions and amino acid residues are more important than others, such that the amino acid network generally represents a subset of all potential

interactions and residues within a protein. In some cases, there may be multiple amino acid networks identified (e.g. ref. [11]), where local changes primarily affect the interactions between the amino acid residues involved in a particular network.

A variety of diverse structural and computational methods have been developed to delineate amino acid networks in proteins, and these methods have provided tremendous insights into protein function. In this Review, we highlight some of the different computational and experimental methods that have been used to delineate amino acid networks in proteins (Table 1), and indicate the insight that these approaches have given regarding protein function. We note that other recent review articles have been written on many of these methods, including graph theory [6], molecular dynamics (MD) simulations [12], elastic network models (ENM [13]), NMR methods to study allostery and amino acid networks [14] and bioinformatics methods to identify co-evolving residues [15], and as such, we do not treat these methods comprehensively. We also recognize that the length of this review prevents us from being exhaustive with our examples.

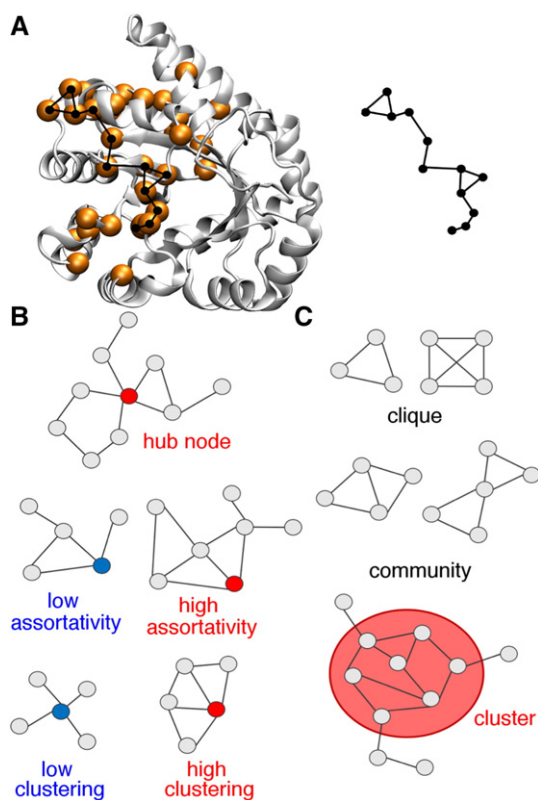
## 2. Network approaches to understanding protein function

In biology, network interactions have been analyzed from the species to the molecular level [16–18]. The elegance of this mathematical theory is to simplify a complex problem into a set of nodes and edges,

<sup>\*</sup> Corresponding author at: 107 Chemistry Building, The Pennsylvania State University, University Park, PA, USA.

E-mail address: [ddb12@psu.edu](mailto:ddb12@psu.edu) (D.D. Boehr).

<sup>1</sup> These authors contributed equally to this work.



**Fig. 1.** Proteins can be viewed as interacting networks of amino acid residues. A. Partial network in the alpha subunit of tryptophan synthase (PDB 1K3U) identified by NMR methods [93]. In the network representation, the nodes are the amino acid residues and represented by circles, and the edges are interactions between the residues and are indicated by lines joining the circles. B. Concepts related to network theory, including hub residues, assortativity and clustering. C. Networks can follow a hierarchy of connectivities, ranging from smaller cliques to larger clusters. Panels B and C were adapted from ref. [38].

together known as a ‘graph’ [19–21]. Graphical approaches have provided intuitive pictures and useful insights for analyzing many complex biological problems, including enzyme-catalyzed reactions [22–24], inhibition of HIV-1 reverse transcriptase [25], inhibition kinetics of processive nucleic acid polymerases and nucleases [26], protein folding kinetics [27] and drug metabolism systems [28]. In the context of protein structure, the amino acid side-chains, or whole amino acid residues, are most commonly treated as the nodes. An edge represents some type of interaction between two nodes. Edges can have a range of definitions such as the calculated energy of interaction, evolutionary conservation, or surface overlap [29–32]. An important feature of edges is the weighting, which may allocate different strengths to different types of interactions and/or provide a particular cut-off distance for residues in close sequence space [33]. There are many algorithms available to construct and analyze amino acid networks using graph theory, including CSU software [34], xPyder [35], PSN-Ensemble [36] and NetworkAnalyzer [37].

Other parameters of the protein graph may be used to further analyze the network, and be related to different structural and functional properties of the protein. For example, a ‘hub node’ has a higher number of edges connected to it than other nodes [38] (Fig. 1). Residues corresponding to hub nodes may be key factors for maintaining structure and determining function. For example, a large experimental set of T4 lysozyme protein variants was studied, where some amino acid substitutions had little to no effect on the function of the enzyme and some substitutions inactivated the protein [39]. All of the deleterious substitutions were later identified as central hub nodes [40].

Connectivity is an important feature of a protein graph. The clustering coefficient,  $C_v$ , provides a measure of connectivity through Eq. (1):

$$C_v = \frac{2e_v}{k_v(k_v-1)} \quad (1)$$

where  $k_v$  is the number of neighbors to node  $v$ , and  $e_v$  is the number of connected pairs among  $v$  neighbors. Residues that have a high connectivity are typically linked to separate clusters or communities of residues [38]. The assortativity matrix is another parameter that helps determine the impact a node has on the network (Fig. 1). This matrix is a measure of the number of connections between nodes. A more ‘resilient’ network [41] would have a higher assortativity, providing multiple paths to connect distant regions of the protein.

A group of nodes can be classified into different types according to how a signal might be transmitted through them (Fig. 1). A clique (or  $k$ -clique) is a complete subgraph, meaning that it is a set of nodes and edges that are connected to every other node in the subgraph [38]. Similar to cliques are communities, which represent a set of connected cliques [38]. Inspection of the cliques and communities in a given protein might be used to track small ligand-induced conformational changes and signal transmission, which can be indicative of the interaction strength of the effector molecule and the quality of the network as a whole. For example, differences in the cliques and communities between the apo and ligand bound states of methionyl t-RNA synthetase were used to understand inter-domain signaling [42]. The binding of ATP induces the formation of new cliques that allow for communication between distal areas of the enzyme.

A cluster has more relaxed requirements than a clique or a community (Fig. 1). In a cluster, the nodes have a higher connectivity with each other than with nodes outside the group, but not all interact pairwise [38]. The largest cluster may be important in defining the core of the protein and can involve up to 80% of the nodes in the entire network. For example, identification of hydrophobic subclusters was used to understand long-range interactions important for stabilizing the tertiary fold of proteins [43]. In this study, it was found that the clusters were larger in thermophilic proteins, which may lead to higher temperature stability.

Measures of residue centrality, including closeness ( $C_n$ ) and betweenness, are often used to predict residues important for the transmission of information across a protein structure. The closeness centrality [44] is defined according to Eq. (2):

$$C_n = \frac{j-1}{\sum_{i \neq n} sd(i, n)} \quad (2)$$

where  $sd(i, n)$  is the shortest path between nodes  $i$  and  $n$ , and  $j$  is the number of nodes in the network. The betweenness centrality  $B$  is determined as the fraction of shortest paths that pass through a node [45]. Residues with high  $C_n$  or  $B$  have been shown to play critical roles in protein function [40,46,47]. Other measures of residue centrality have been proposed (e.g. ref. [48,49] and references therein). Recent examples using these types of approaches include studies analyzing allosteric pathways in tRNA synthetases [50], G-protein coupled receptors [51], Hsp90 [52] and cyclophilin A [53].

### 3. More sophisticated structure-based approaches to network analysis

Conformational fluctuations in proteins are important in mediating their biological functions. For example, *E. coli* dihydrofolate reductase (DHFR) must pass through multiple conformations as it proceeds through its catalytic cycle [54]. Smaller fluctuations, such as those in side chains, may be evident in X-ray diffraction data [55], though they may be ignored during the refinement process when producing a structural model. The qFit algorithm was developed to fit these alternative

**Table 1**

Brief summary of the computational and biophysical methods to analyze amino acid networks in proteins addressed in this review.

	Brief description	Comments
Graph theory	Formalized math-based approach to identify residue connectivities and potential allosteric paths in proteins.	Often used with static protein structures, which may not reveal the full range of potential contacts in a dynamic protein. Combining graph theory and dynamic information (e.g. through MD simulations) is a powerful approach.
CONTACT	Identify alternative side-chain conformations and their contacts based on high quality X-ray diffraction data and electron density maps.	High quality diffraction data is generally required, especially at non-cryogenic temperatures. This method can be combined with principles from graph theory.
MD simulations and elastic network approaches	Computer simulations and calculations of internal protein motion over a wide range of timescales.	These computer simulations offer trajectories of motions, often missing in biophysical approaches that study protein structural dynamics. These methods are often combined with principles from graph theory. Non-equilibrium approaches offer additional ways to identify potential allosteric pathways. These approaches become more convincing with experimental validation.
Analysis of amino acid perturbations using NMR	Perturbations of protein structure/dynamic (e.g. through amino acid substitutions) can be followed on an atom-specific basis to reveal potential allosteric pathways.	NMR offers great experimental methods to analyze internal motions of proteins over a wide range of timescales at atomic resolution. These methods are often used to validate MD simulations, and/or extend analysis to longer timescales (i.e. >microseconds). Solution-state NMR may be limited to smaller proteins (i.e. < 100 kDa).
Amino acid sequence based analysis of networks	In an amino acid sequence alignment of hundreds to thousands of similar proteins, identify residue positions that covary, implying coevolution.	An appropriate number of protein sequences and an appropriate level of sequence diversity are required. The information may be difficult to understand in the absence of a structural/dynamic rationale.

side-chain conformations into the electron density derived from diffraction data [56]. The qFit algorithm is especially useful for analyzing high quality diffraction data that was collected at ambient temperature, where more substantial protein motions within the crystal would be expected [57]. These alternate side-chain conformations may lead to different interactions, providing additional information in crafting a network.

The CONTACT algorithm has been developed to analyze potential effects of alternative side-chain conformations [58]. For the networks identified by CONTACT, nodes are defined as the side-chain and edges are defined as the steric clash of van der Waals radii between residues with alternate conformations. In the CONTACT algorithm, once it has been determined that there is a steric clash present between residues *i* and *j*, the algorithm switches residue *j* to one of its alternative conformations (Fig. 2). If this change results in another clash with some residue *k* then the algorithm moves *k* to a different conformation. This process repeats until all steric clashes are relieved, thereby identifying a particular pathway. CONTACT is capable of identifying a variety of different paths in a single network, so the edges are weighted by the number of pathways in which residues *i* and *j* are predicted to clash. This method was used to generate networks for cyclophilin A [59] and *E. coli* DHFR [58]; amino acid substitutions at CONTACT-determined network positions led to dramatic decreases in enzyme activity in both of these systems.

#### 4. The importance of internal motions to amino acid networks

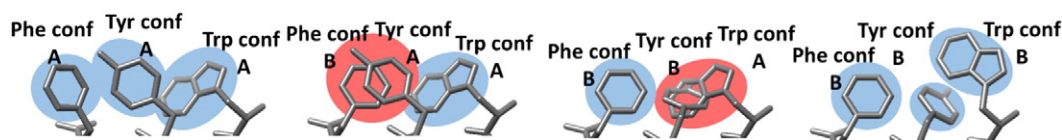
The network approaches described above primarily focus on a limited set of protein conformations. It is now recognized that rarely visited conformations can have important effects on protein function [60]. Such lowly populated conformations would have different sets of noncovalent interactions, and thus, potentially different amino acid networks. Computer simulations offer a means to predict and analyze these different amino acid networks. Such simulations can be categorized into molecular dynamics (MD) simulations and more coarse-grained models.

#### 4.1. MD simulations

It has been recognized for decades that protein structure is dynamic [61,62]. Low frequency, collective motions have been shown to exist in proteins and nucleic acids [63–65], and these motions can be important for a variety of protein functions, including switching between active and inactive states [66], cooperative effects [67], allosteric transitions [68] and assembly of microtubules [69]. In view of these important properties of proteins, it is imperative to consider not only static structural information but also the internal motions of proteins; MD simulations offer one way to do this.

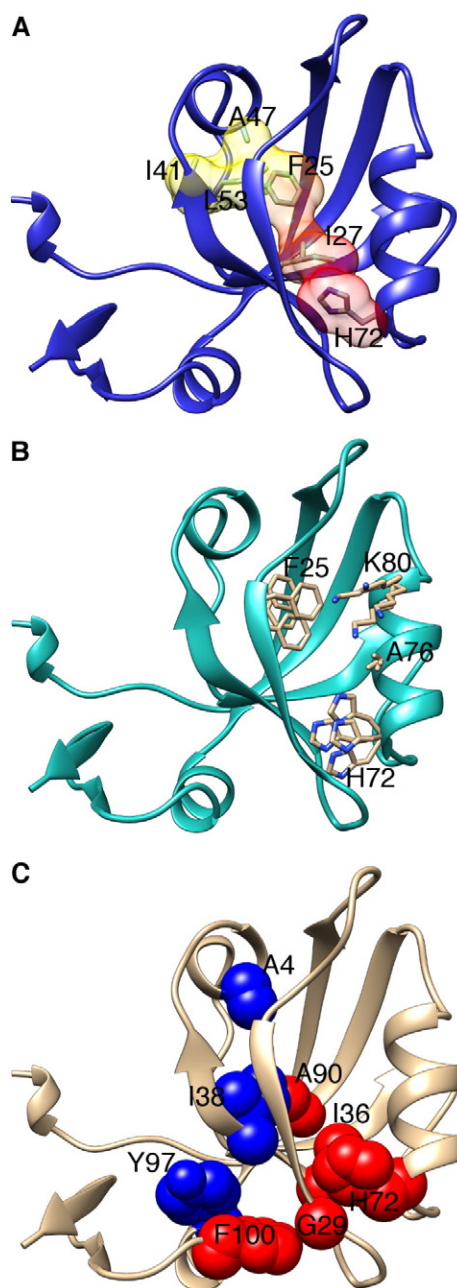
MD simulations can provide trajectories of atoms or residues within a protein. One common approach to analyze these trajectories is to quantify correlated or anti-correlated motions using the dynamic cross-correlation map [70]; such motions may provide information about more concerted motions important for protein function. For example, Moustafa et al. [71] showed that the G64S substitution in the poliovirus RNA-dependent RNA polymerase leads to a higher fidelity polymerase [72], perhaps by changing (anti)correlated motions in conserved regions of the enzyme.

Non-equilibrium perturbation methods compatible with MD have also provided insight into how signals might be propagated through a protein. These methods include anisotropic thermal diffusion (ATD) [73], pump-probe molecular dynamics (PPMD) [74], and rigid residue scan (RRS) [75]. ATD is a method that examines the transfer of thermal energy through non-bonding interactions, primarily through van der Waals interactions between sidechains, by cooling the system to 10 K and applying thermal energy to a residue known to be important for function [73,75]. PPMD reveals networks by applying an oscillating motion of varying frequency to an  $\alpha$ -carbon, or more atoms at a greater computational cost, and identifying residues that couple with this perturbation over a nanosecond timescale [74]. An advantage of the PPMD method is that it can be applied to any standard MD simulation and is compatible with most force fields [74]. RRS creates networks by treating specific residues as rigid bodies, and analyzes how this treatment affects the conformational properties of the protein [75].



**Fig. 2.** Alternate side-chain conformations using the CONTACT algorithm can be used to generate a network. For example, when the Phe sidechain transitions to conformation B there is a clash in the van der Waals radii with the Tyr sidechain. This clash is alleviated by switching Tyr to conformation B, which induces another clash which is then relieved by changing the Trp conformation. This process is repeated until there are no more steric clashes. This figure was adapted from Ref. [58].



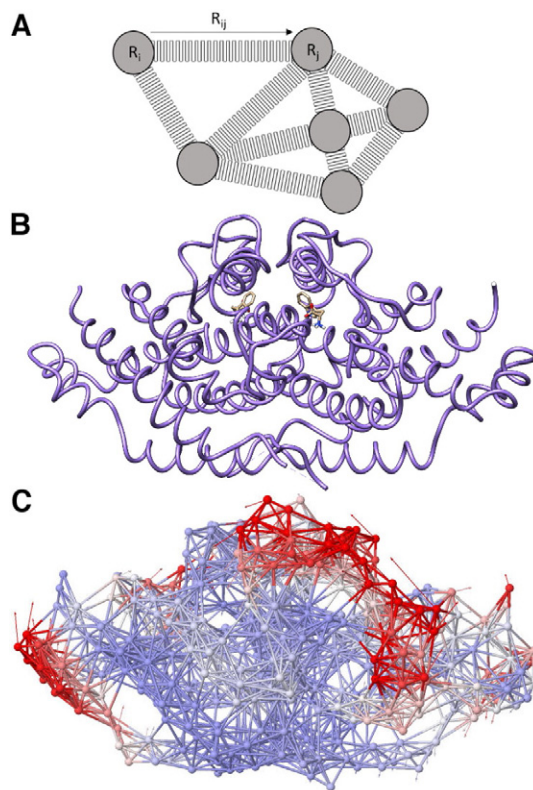


**Fig. 3.** Nonequilibrium perturbations methods have been used to study amino acid networks in the PDZ domain. A. Anisotropic thermal diffusion simulations start by cooling the system to 10 K, followed by the application of a 300 K heat bath to a residue of interest. In this case, His72, an active site residue, was heated and thermal energy was diffused through van der Waals interactions Ile27, Phe25, Ile41, Ala47, and Leu53 [73]. B. Pump probe molecular dynamics consist of the application of oscillating motions at varying frequencies and directions to an atom or set of atoms and observing the transfer of that motion to other residues. The oscillations at His72 transferred to Ala76, Lys80, and Phe25. Different oscillation frequencies and directions tend to show coupling to different residues, and the above representation does not represent all coupling interactions observed [74]. C. Rigid residue scan was used to systematically apply rigid body constraints to each residue in the PDZ domain to simulate point mutations [75]. By utilizing heat maps to look at the difference between the ligand-bound and unbound state of the domain during a simulation with SHAPE constraints on each residue, some residues were identified as switches (blue, including residues Ile8, Ala47, and Tyr97) while other important residues were designated as wires (red, residues Gly29, Ile36, His72, Ala90, and Phe100) depending on the degree to which they effect the transition between the bound and unbound state. Panels A, B and C were adapted from references [73–75], using PDB 1BE9.

These methods were applied to understanding the allosteric transition of the PDZ3 protein [73–75] (Fig. 3). RRS identified a network of nine residues that was deemed important for energy propagation through the PDZ3 domain [75]. Switch residues were identified as those that when rigidified in the simulation disrupted the conformational change, whereas residues that had smaller effects on the conformational change were designated as wires [75] (Fig. 3).

#### 4.2. Elastic network models

Course-grain simulations often employ normal mode analysis (NMA), which simulates the harmonic motions of a system. When performed on proteins, the lower frequency modes provide information about potential concerted, global motions. Early NMA was limited to smaller systems due to the computational cost of simulating every interaction between every atom. More modern applications of NMA typically take advantage of a coarse graining technique known as the elastic network model (ENM), which has allowed for the analysis of large macromolecular complexes such as the GroEL tetradecamer [76–78]. ENM simulations have such low hardware requirements compared to MD simulations that they can be simulated using web-based tools such as ANM 2.0 [79] (e.g. see Fig. 4). Modifications and improvements to ENM include the Gaussian network model (GNM) [80], the anisotropic network method (ANM) [81], the robust elastic network model (RENM) and the heterogeneous anisotropic elastic network model (HANM) [82,83].



**Fig. 4.** The elastic network model simulates global harmonic motions by simplifying biomolecules to a network of interacting nodes. A. In these types of models,  $R_i$  and  $R_j$  are interacting residues, and their interactions create a separation vector that is represented by  $R_{ij}$ . The separation vector can be essentially treated as a 'spring' with a given spring constant, depending on the interaction energy. The yeast chorismate mutase homodimer (PDB 1CSM) with tryptophan bound is shown as: B. a ribbon diagram, and C. a network representation with the help of ANM 2.0, a web-based tool for ENM normal mode simulations [79]. The network representation shown is 'mode 16', with the arrows indicating the direction of motion. ENM typically treats each alpha carbon as a node in a network of springs and nodes. The springs represent interactions within a set cutoff distance.

As an example, NMA was applied for an ENM-based simulation of the maltose transporter MalFGK2 [84]. Three low-frequency modes that contribute to channel-gating motions were analyzed, and residues involved in these low-frequency modes were perturbed by changing the spring constant between the residue and surrounding residues. The perturbations that created the greatest change in normal mode frequencies were mapped onto the protein, and the identified residues formed a network that was found to connect the channel gate on the periplasmic side of the complex, the cytoplasmic side, and the interaction points of the subunits of the complex [11].

A similar perturbation method, fluctuation perturbation analysis (FPA), was applied to a variety of proteins by Zheng et al. [85]. For example, FPA was performed to predict key residues involved in the conformational change of myosin during its power stroke, which is a necessary part of muscle contraction. The residues formed a 3D network connecting the active site to the force-generating component and the actin binding cleft [85].

## 5. NMR perturbation methods

NMR studies provide insight into the conformational dynamics of proteins over a wide-range of timescales [86]. One way to gain information about potential networks is to perturb the protein in some way (e.g. ligand binding, amino acid substitutions) and analyze correlated changes to NMR parameters. For example, Clarkson et al. generated a series of Val to Ala substitutions in eglin c and mapped dynamic responses by NMR experiments across different timescales to identify networks that propagate primarily through structural or dynamic changes [87]. The same research group performed a similar analysis with the PDZ domain [88], whose results were consistent with the MD methods discussed above.

Correlations between chemical shift changes induced by a series of perturbations can also help define a network [31,89]. Perturbations of the protein conformational equilibrium through ligand binding or amino acid substitutions can result in chemical shift changes both local and remote from the site(s) of perturbation. Residues with covarying chemical shift changes across a series of perturbations are proposed to be involved in the same conformational change and thus belong to the same amino acid network [31]. The original method was developed

by Melacini and colleagues [31], and was termed CHESCA (CHEMical Shift Covariance Analysis). Melacini and coworkers used their algorithm to define two different amino acid networks in the protein EPAC; one of these networks was associated with ligand binding and the other network was thought to be important for allosteric signaling [31]. Other groups have developed similar algorithms (e.g. Ref. [90,91]).

In our own work, we have used a method similar to CHESCA to define amino acid networks in the alpha subunit of tryptophan synthase [92,93] (Fig. 5). In our approach, we used a series of amino acid substitutions as our source of perturbations, which allowed us to interrogate amino acid networks in the ligand-free protein (i.e. the 'resting' state) and when the enzyme was actively turning over (i.e. the 'working' state) [93]. Remarkably, the amino acid networks were different between the resting and working states [93]. Of special importance was the behavior of the catalytically-important residue Glu49, which was allocated to different clusters in the resting versus working states [93]. Previous work had also shown that Glu49 undergoes an important conformational change upon binding substrate [94]; the change in network interactions may help to re-position Glu49 for catalysis.

## 6. Bioinformatic approaches to delineating amino acid networks

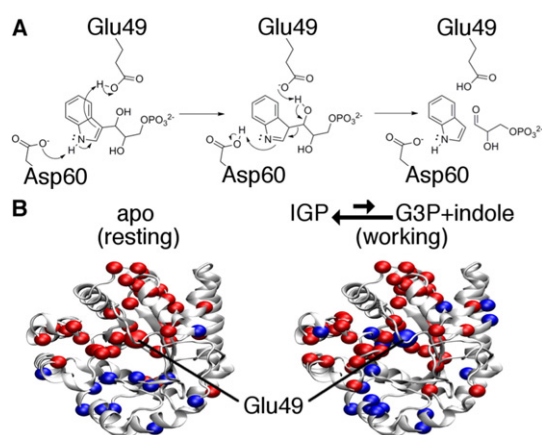
Since the amino acid networks are functionally-important then they may also be evolutionarily conserved. Insight into amino acid networks might then be provided by analysis of multiple sequence alignments (MSA) in the absence of structural information, as has been done through algorithms including statistical coupling analysis (SCA [10,21,95]), mutual information (MI [96]), McLachlan Based Substitution Correlation (McBASC [97]) and Observed Minus Expected Square (OMES [98]). A more thorough evaluation of these methods is found in Livesay et al. (2012) [15].

The SCA method has been successfully used to identify functionally significant networks in the PDZ domain [99], plant peroxidase [100], Fe/Mn superoxide dismutase [101], G protein-coupled receptor, chymotrypsin class serine protease, and hemoglobin families [102]. In SCA, energetic connectivity between amino acid residues is taken as being evolutionarily conserved to bestow common functions within protein families [99]. A key assumption of the SCA method is that the probability of finding a particular residue at a particular position without any evolutionary constraints on that position will be the same as the mean abundance of that residue in all proteins [99]; deviation from that mean frequency indicates conservation. The co-conservation of residues at two or more positions in the MSA indicates statistical coupling [99]. The statistical degree to which two sites are coupled is measured by the frequency that a change in the amino acid identity at one site is associated with a change in the coupled site [99]. The magnitude of the evolutionary constraint on a particular position is described through a Boltzmann distribution [99].

One particular noteworthy example of the power of the SCA approach is found in the engineering of a light-responsive dihydrofolate reductase (DHFR) enzyme. Ranganathan et al. used SCA on an MSA of 418 sequences to provide a basis for the discovery of novel allostery in DHFR [103]. SCA revealed a 'sector' that acts as a wire, which transmits allosteric signals between surface residues and the active site. To further investigate how these wires might affect DHFR activity, a light-sensitive PAS domain, LOV2, was inserted into the DHFR sequence at solvent-exposed sites. This library of DHFR-LOV2 chimeras was characterized through kinetic studies, and it was found that the chimeras with the LOV2 domain inserted at sites that were part of the cluster had an altered catalytic efficiency as compared to those that had the LOV2 domain inserted at non-cluster sites [103].

## 7. Summary and outlook

The characterization of amino acid networks by diverse experimental and computational methods has provided deeper understanding of



**Fig. 5.** Amino acid networks in the alpha subunit of *E. coli* tryptophan synthase ( $\alpha$ TS) are dependent on what is bound to the enzyme. A. Conversion of indole-3-glycerol phosphate (IGP) to glyceraldehyde-3-phosphate (G3P) and indole catalyzed by  $\alpha$ TS, highlighting the roles of Glu49 and Asp60 in the chemical mechanism. B. The NMR method CHESCA [31] was used to delineate amino acid networks in both the 'resting state' in the absence of ligands (left) and 'working state' during active turnover [93]. In both cases, two clusters, represented by red and blue spheres, were identified. Intriguingly, the catalytic residue Glu49 changes clusters from the resting to working states, implying that these clusters might be important in regulating the catalytic activity of  $\alpha$ TS. The protein structure is derived from PDB 1K3U.



the function and regulation of proteins. Network analyses have provided insight into internal pathways that may be important for propagating allosteric signals, and have suggested that all proteins have intrinsic allosteric properties [104]. Such properties would help to explain how amino acid changes remote from a ligand binding-site or enzyme active site [105] can nonetheless influence the function of the protein. This view has practical consequences, including in understanding drug resistance (e.g. ref. [106]), the development of new allosteric drugs that may target surface-exposed network residues [107,108], and in protein engineering applications [109,110], where such proteins would have wide ranging applications from biosensing to synthetic biology. The integration of various methods (e.g. ref. [111]) will likely provide greater insight into amino acid networks and provide additional lessons regarding the strengths and weaknesses of various approaches. These approaches will continue to increase our understanding of protein function, and provide novel avenues towards modulating these functions in practical applications of drug and protein design.

## Acknowledgements

We thank Alyson Boehr, Rebecca D'Amico and Dr. Ibrahim Moustafa for their comments on the current manuscript and for conversations regarding amino acid networks. The work in the Boehr laboratory on amino acid networks in proteins is supported by NSF Career Grant MCB1053993 and NIH grant R01AI104878.

## References

- [1] Chou KC, Nemethy G, Scheraga HA. Energetics of interactions of regular structural elements in proteins. *Acc Chem Res* 1990;23:134–41.
- [2] Chou KC, Maggiora GM, Scheraga HA. Role of loop–helix interactions in stabilizing four-helix bundle proteins. *Proc Natl Acad Sci U S A* 1992;89:7315–9.
- [3] Du QS, Chen D, Xie NZ, Huang RB, Chou KC. Insight into a molecular interaction force supporting peptide backbones and its implication to protein loops and folding. *J Biomol Struct Dyn* 2015;33:1957–72.
- [4] Grewal RK, Roy S. Modeling proteins as residue interaction networks. *Protein Pept Lett* 2015;22:923–33.
- [5] Bhattacharyya M, Ghosh S, Vishveshwara S. Protein structure and function: looking through the network of side-chain interactions. *Curr Protein Pept Sci* 2016;17:4–25.
- [6] Di Paola L, Giuliani A. Protein contact network topology: a natural language for allostery. *Curr Opin Struct Biol* 2015;31:43–8.
- [7] Rivalta I, Sultan MM, Lee NS, Manley GA, Loria JP, Batista VS. Allosteric pathways in imidazole glycerol phosphate synthase. *Proc Natl Acad Sci U S A* 2012;109:E1428–36.
- [8] Datta D, Scheer JM, Romanowski MJ, Wells JA. An allosteric circuit in caspase-1. *J Mol Biol* 2008;381:1157–67.
- [9] 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.
- [10] Halabi N, Rivoire O, Leibler S, Ranganathan R. Protein sectors: evolutionary units of three-dimensional structure. *Cell* 2009;138:774–86.
- [11] Gasper PM, Fuglestad B, Komives EA, Markwick PR, McCammon JA. Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities. *Proc Natl Acad Sci U S A* 2012;109:21216–22.
- [12] Feher VA, Durrant JD, Van Wart AT, Amaro RE. Computational approaches to mapping allosteric pathways. *Curr Opin Struct Biol* 2014;25:98–103.
- [13] Fuglebakk E, Tiwari SP, Reuter N. Comparing the intrinsic dynamics of multiple protein structures using elastic network models. *Biochim Biophys Acta* 2015;1850:911–22.
- [14] Grutsch S, Bruschweiler S, Tollinger M. NMR methods to study dynamic allostery. *PLoS Comput Biol* 2016;12, e1004620.
- [15] Livesay DR, Kretz KE, Fodor AA. A critical evaluation of correlated mutation algorithms and coevolution within allosteric mechanisms. *Methods Mol Biol* 2012;796:385–98.
- [16] Atilgan C, Okan OB, Atilgan AR. Network-based models as tools hinting at nonevident protein functionality. *Annu Rev Biophys* 2012;41:205–25.
- [17] Raman K. Construction and analysis of protein–protein interaction networks. *Autom Exp* 2010;2:2.
- [18] Vidal M, Cusick ME, Barabási A-L. Interactome networks and human disease. *Cell* 2011;144:986–98.
- [19] Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D* 2004;60:2256–68.
- [20] Przytycka T, Srinivasan R, Rose GD. Recursive domains in proteins. *Protein Sci* 2002;11:409–17.
- [21] Suel GM, Lockless SW, Wall MA, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol* 2003;10:59–69.
- [22] Chou KC, Forsen S. Graphical rules for enzyme-catalysed rate laws. *Biochem J* 1980;187:829–35.
- [23] Zhou GP, Deng MH. An extension of Chou's graphic rules for deriving enzyme kinetic equations to systems involving parallel reaction pathways. *Biochem J* 1984;222:169–76.
- [24] Chou KC. Graphic rules in steady and non-steady state enzyme kinetics. *J Biol Chem* 1989;264:12074–9.
- [25] Althaus IW, Gonzales AJ, Chou JJ, Romero DL, Deibel MR, Chou KC, et al. The quinoline U-78036 is a potent inhibitor of HIV-1 reverse transcriptase. *J Biol Chem* 1993;268:14875–80.
- [26] Chou KC, Kezdy FJ, Reusser F. Kinetics of processive nucleic acid polymerases and nucleases. *Anal Biochem* 1994;221:217–30.
- [27] Chou KC. Applications of graph theory to enzyme kinetics and protein folding kinetics. Steady and non-steady-state systems. *Biophys Chem* 1990;35:1–24.
- [28] Chou KC. Graphic rule for drug metabolism systems. *Curr Drug Metab* 2010;11:369–78.
- [29] Sethi A, Eargle J, Black AA, Luthey-Schulten Z. Dynamical networks in tRNA:protein complexes. *Proc Natl Acad Sci U S A* 2009;106:6620–5.
- [30] Bhattacharyya M, Upadhyay R, Vishveshwara S. Interaction signatures stabilizing the NAD(P)-binding Rossmann fold: a structure network approach. *PLoS One* 2012;7, e51676.
- [31] 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;108:6133–8.
- [32] Basu S, Bhattacharyya D, Banerjee R. Mapping the distribution of packing topologies within protein interiors shows predominant preference for specific packing motifs. *BMC Bioinformatics* 2011;12:1–26.
- [33] Leelananda SP, Towfic F, Jernigan RL, Kloczkowski A. Exploration of the relationship between topology and designability of conformations. *J Chem Phys* 2011;134:235101.
- [34] Sobolev V, Sorokine A, Prilusky J, Abola EE, Edelman M. Automated analysis of interatomic contacts in proteins. *Bioinformatics* 1999;15:327–32.
- [35] Pasi M, Tiberti M, Arrigoni A, Papaleo E. xPyder: a PyMOL plugin to analyze coupled residues and their networks in protein structures. *J Chem Inf Model* 2012;52:1865–74.
- [36] Bhattacharyya M, Bhat CR, Vishveshwara S. An automated approach to network features of protein structure ensembles. *Protein Sci* 2013;22:1399–416.
- [37] Doncheva NT, Assenov Y, Domingues FS, Albrecht M. Topological analysis and interactive visualization of biological networks and protein structures. *Nat Protoc* 2012;7:670–85.
- [38] Vishveshwara S, Brinda KV, Kannan N. Protein structure: insights from graph theory. *J Theor Comput Chem* 2002;01:187–211.
- [39] Rennell D, Bouvier SE, Hardy LW, Poteete AR. Systematic mutation of bacteriophage T4 lysozyme. *J Mol Biol* 1991;222:67–88.
- [40] Amitai G, Shemesh A, Sitbon E, Shklar M, Netanel D, Venger I, et al. Network analysis of protein structures identifies functional residues. *J Mol Biol* 2004;344:1135–46.
- [41] Newman MEJ, Strogatz SH, Watts DJ. Random graphs with arbitrary degree distributions and their applications. *Phys Rev E* 2001;64:026118.
- [42] Ghosh A, Vishveshwara S. Variations in clique and community patterns in protein structures during allosteric communication: investigation of dynamically equilibrated structures of methionyl tRNA synthetase complexes. *Biochemistry* 2008;47:11398–407.
- [43] Sengupta D, Kundu S. Role of long- and short-range hydrophobic, hydrophilic and charged residues contact network in protein's structural organization. *BMC Bioinformatics* 2012;13:142.
- [44] Sabidussi G. The centrality of a graph. *Psychometrika* 1966;31:581–603.
- [45] Freeman LC. Set of measures of centrality based on betweenness. *Sociometry* 1977;40:35–41.
- [46] Di Paola L, De Ruvo M, Paci P, Santoni D, Giuliani A. Protein contact networks: an emerging paradigm in chemistry. *Chem Rev* 2013;113:1598–613.
- [47] del Sol A, Fujihashi H, O'Meara P. Topology of small-world networks of protein–protein complex structures. *Bioinformatics* 2005;21:1311–5.
- [48] Karain WI, Qaraeen NI. Weighted protein residue networks based on joint recurrences between residues. *BMC Bioinformatics* 2015;16:173.
- [49] Karabekmez ME, Kirdar B. A novel topological centrality measure capturing biologically important proteins. *Mol Biosyst* 2016;12:666–73.
- [50] Bhattacharyya M, Vishveshwara S. Probing the allosteric mechanism in pyrrolysyl-tRNA synthetase using energy-weighted network formalism. *Biochemistry* 2011;50:6225–36.
- [51] Lee Y, Choi S, Hyeon C. Mapping the intramolecular signal transduction of G-protein coupled receptors. *Proteins* 2014;82:727–43.
- [52] Blacklock K, Verkhivker GM. 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.
- [53] Doshi U, Holliday MJ, Eisenmesser EZ, Hamelberg D. Dynamical network of residue-residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. *Proc Natl Acad Sci U S A* 2016;113:4735–40.
- [54] Boehr DD, McElheny D, Dyson HJ, Wright PE. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* 2006;313:1638–42.
- [55] Terwilliger TC, Grosse-Kunstele RW, Afonine PV, Adams PD, Moriarty NW, Zwart P, et al. Interpretation of ensembles created by multiple iterative rebuilding of macromolecular models. *Acta Crystallogr D Biol Crystallogr* 2007;63:597–610.
- [56] van den Bedem H, Dhanik A, Latombe J-C, Deacon AM. Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers. *Acta Crystallogr D Biol Crystallogr* 2009;65:1107–17.

- [57] Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, et al. Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc Natl Acad Sci U S A* 2011;108:16247–52.
- [58] van den Bedem H, Bhabha G, Yang K, Wright PE, Fraser JS. Automated identification of functional dynamic contact networks from X-ray crystallography. *Nat Methods* 2013;10:896–902.
- [59] Fraser JS, Clarkson MW, Degnan SC, Erion R, Kern D, Alber T. Hidden alternative structures of proline isomerase essential for catalysis. *Nature* 2009;462:669–73.
- [60] Kay LE. New views of functionally dynamic proteins by solution NMR spectroscopy. *J Mol Biol* 2016;428:323–31.
- [61] Chou KC, Chen NY. Biological functions of low-frequency phonons. *Sci Sin* 1977;20:447–57.
- [62] Frauenfelder H, Sligar SG, Wolynes PG. The energy landscapes and motions of proteins. *Science* 1991;254:1598–603.
- [63] Chou KC, Maggiora GM, Mao B. Quasi-continuum models of twist-like and accordion-like low-frequency motions in DNA. *Biophys J* 1989;56:295–305.
- [64] Martel P. Biophysical aspects of neutron scattering from vibrational modes of proteins. *Prog Biophys Mol Biol* 1992;57:129–79.
- [65] Chou KC. Low-frequency collective motion in biomacromolecules and its biological functions. *Biophys Chem* 1988;30:3–48.
- [66] Wang JF, Gong K, Wei DQ, Li YX, Chou KC. Molecular dynamics studies on the interactions of PTP1B with inhibitors: from the first phosphate-binding site to the second one. *Protein Eng Des Sel* 2009;22:349–55.
- [67] Chou KC. Low-frequency resonance and cooperativity of hemoglobin. *Trends Biochem Sci* 1989;14:212–3.
- [68] Chou KC, Chen NY, Forsen S. The biological functions of low-frequency phonons 2. Cooperative effects. *Chem Scr* 1981;18:126–32.
- [69] Chou KC, Zhang CT, Maggiora GM. Solitary wave dynamics as a mechanism for explaining the internal motion during microtubule growth. *Biopolymers* 1994;34:143–53.
- [70] Ichiye T, Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* 1991;11:205–17.
- [71] Moustafa IM, Shen H, Morton B, Colina CM, Cameron CE. Molecular dynamics simulations of viral RNA polymerases link conserved and correlated motions of functional elements to fidelity. *J Mol Biol* 2011;410:159–81.
- [72] Arnold JJ, Vignuzzi M, Stone JK, Andino R, Cameron CE. Remote site control of an active site fidelity checkpoint in a viral RNA-dependent RNA polymerase. *J Biol Chem* 2005;280:25706–16.
- [73] Ota N, Agard DA. Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. *J Mol Biol* 2005;351:345–54.
- [74] Sharp K, John JS. Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling. *Proteins* 2006;65:347–61.
- [75] Kalescky R, Liu J, Tao P. Identifying key residues for protein allostery through rigid residue scan. *J Phys Chem A* 2014.
- [76] Na H, Jernigan RL, Song G. Bridging between NMA and elastic network models: preserving all-atom accuracy in coarse-grained models. *PLoS Comput Biol* 2015;11:e1004542.
- [77] Na H, Song G. Bridging between normal mode analysis and elastic network models. *Proteins* 2014;1–12.
- [78] Skjaerven L, Martinez A, Reuter N. Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit. *Proteins Struct Funct Bioinformatics* 2011;79:232–43.
- [79] Eyal E, Lum G, Bahar I. The anisotropic network model web server at 2015 (ANM 2.0). *Bioinformatics* 2015;31:1487–9.
- [80] Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* 1997;2:173–81.
- [81] Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J* 2001;80:505–15.
- [82] Kim MH, Lee BH, Kim MK. Robust elastic network model: a general modeling for precise understanding of protein dynamics. *J Struct Biol* 2015;190:338–47.
- [83] Li M, Zhang JZH, Xia F. Heterogeneous elastic network model improves description of slow motions of proteins in solution. *Chem Phys Lett* 2015;618:102–7.
- [84] Su JG, Han XM, Zhang X, Hou YX, Zhu JZ, Wu YD. Analysis of conformational motions and related key residue interactions responsible for a specific function of proteins with elastic network model. *J Biomol Struct Dyn* 2015;1102:1–12.
- [85] Zheng W, Tekpinar M. Large-scale evaluation of dynamically important residues in proteins predicted by the perturbation analysis of a coarse-grained elastic model. *BMC Struct Biol* 2009;9:45.
- [86] Palmer III AG. Chemical exchange in biomacromolecules: past, present, and future. *J Magn Reson* 2014;241:3–17.
- [87] Clarkson MW, Gilmore SA, Edgell MH, Lee AL. Dynamic coupling and allosteric behavior in a nonallosteric protein. *Biochemistry* 2006;45:7693–9.
- [88] Fuentes EJ, Gilmore SA, Mauldin RV, Lee AL. Evaluation of energetic and dynamic coupling networks in a PDZ domain protein. *J Mol Biol* 2006;364:337–51.
- [89] Boulton S, Akimoto M, Selvaratnam R, Bashiri A, Melacini G. A tool set to map allosteric networks through the NMR chemical shift covariance analysis. *Sci Rep* 2014;4:7306.
- [90] Cembran A, Kim J, Gao J, Veglia G. NMR mapping of protein conformational landscapes using coordinated behavior of chemical shifts upon ligand binding. *Phys Chem Chem Phys* 2014;16:6508–18.
- [91] Dawson JE, Farber PJ, Forman-Kay JD. Allosteric coupling between the intracellular coupling helix 4 and regulatory sites of the first nucleotide-binding domain of CFTR. *PLoS One* 2013;8:e74347.
- [92] Axe JM, O'Rourke KF, Kerstetter NE, Yezdimer EM, Chan YM, Chasin A, et al. Severing of a hydrogen bond disrupts amino acid networks in the catalytically active state of the alpha subunit of tryptophan synthase. *Protein Sci* 2015;24:484–94.
- [93] Axe JM, Yezdimer EM, O'Rourke KF, Kerstetter NE, You W, Chang CE, et al. Amino acid networks in a (beta/alpha)(8) barrel enzyme change during catalytic turnover. *J Am Chem Soc* 2014;136:6818–21.
- [94] Rhee S, Miles EW, Davies DR. Cryo-crystallography of a true substrate, indole-3-glycerol phosphate, bound to a mutant (alphaD60N) tryptophan synthase alpha2beta2 complex reveals the correct orientation of active site alphaGlu49. *J Biol Chem* 1998;273:8553–5.
- [95] Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* 1999;286:295–9.
- [96] Clarke ND. Covariation of residues in the homeodomain sequence family. *Protein Sci* 1995;4:2269–78.
- [97] McLachlan AD. Tests for comparing related amino-acid sequences. *Cytochrome c and cytochrome c 551. J Mol Biol* 1971;61:409–24.
- [98] Kass I, Horovitz A. Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations. *Proteins* 2002;48:611–7.
- [99] Lockless SW, Ranganathan R, Kukic P, Mirabello C, Tradigo G, Walsh I, et al. Evolutionarily conserved pathways of energetic connectivity in protein families. *BMC Bioinformatics* 1999;15:295–9.
- [100] Watanabe L, de Moura PR, Bleicher L, Nascimento AS, Zamorano LS, Calvete JJ, et al. Crystal structure and statistical coupling analysis of highly glycosylated peroxidase from royal palm tree (*Roystonea regia*). *J Struct Biol* 2010;169:226–42.
- [101] Bachege JFR, Navarro M, Bleicher L, Bortoleto-Bugs RK, Dive D, Hoffmann P, et al. Systematic structural studies of iron superoxide dismutases from human parasites and a statistical coupling analysis of metal binding specificity. *Proteins Struct Funct Bioinformatics* 2009;77:26–37.
- [102] Süel GM, Lockless SW, Wall Ma, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol* 2003;10:59–69.
- [103] Reynolds KA, McLaughlin RN, Ranganathan R. Hot spots for allosteric regulation on protein surfaces. *Cell* 2011;147:1564–75.
- [104] Gunasekaran K, Ma B, Nussinov R. Is allostery an intrinsic property of all dynamic proteins? *Proteins* 2004;57:433–43.
- [105] Lee J, Goodey NM. Catalytic contributions from remote regions of enzyme structure. *Chem Rev* 2011;111:7595–624.
- [106] Appadurai R, Senapati S. Dynamical network of HIV-1 protease mutants reveals the mechanism of drug resistance and unhindered activity. *Biochemistry* 2016;55:1529–40.
- [107] Guarnera E, Berezovsky IN. Allosteric sites: remote control in regulation of protein activity. *Curr Opin Struct Biol* 2015;37:1–8.
- [108] Nussinov R, Tsai CJ, Csermely P. Allo-network drugs: harnessing allostery in cellular networks. *Trends Pharmacol Sci* 2011;32:686–93.
- [109] Bhabha G, Biel JT, Fraser JS. Keep on moving: discovering and perturbing the conformational dynamics of enzymes. *Acc Chem Res* 2015;48:423–30.
- [110] Osuna S, Jimenez-Oses G, Noey EL, Houk KN. Molecular dynamics explorations of active site structure in designed and evolved enzymes. *Acc Chem Res* 2015;48:1080–9.
- [111] Manley G, Rivalta I, Loria JP. Solution NMR and computational methods for understanding protein allostery. *J Phys Chem B* 2013;117:3063–73.