

Research Article

A self-consistent structural perturbation approach for determining the magnitude and extent of allosteric coupling in proteins

Nandakumar Rajasekaran and Athi N. Naganathan

Department of Biotechnology, Bhupat & Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India

Correspondence: Athi N. Naganathan (athi@iitm.ac.in)

Elucidating the extent of energetic coupling between residues in single-domain proteins, which is a fundamental determinant of allostery, information transfer and folding cooperativity, has remained a grand challenge. While several sequence- and structure-based approaches have been proposed, a self-consistent description that is simultaneously compatible with unfolding thermodynamics is lacking. We recently developed a simple structural perturbation protocol that captures the changes in thermodynamic stabilities induced by point mutations within the protein interior. Here, we show that a fundamental residue-specific component of this perturbation approach, the coupling distance, is uniquely sensitive to the environment of a residue in the protein to a distance of ~ 15 Å. With just the protein contact map as an input, we reproduce the extent of percolation of perturbations within the structure as observed in network analysis of intra-protein interactions, molecular dynamics simulations and NMR-observed changes in chemical shifts. Using this rapid protocol that relies on a single structure, we explain the results of statistical coupling analysis (SCA) that requires hundreds of sequences to identify functionally critical sectors, the propagation and dissipation of perturbations within proteins and the higher-order couplings deduced from detailed NMR experiments. Our results thus shed light on the possible mechanistic origins of signaling through the interaction network within proteins, the likely distance dependence of perturbations induced by ligands and post-translational modifications and the origins of folding cooperativity through many-body interactions.

Introduction

Signal transmission is at the core of chemical reactions that drive biological processes in the cellular milieu, not just through the protein–protein interaction network but also within a single protein domain [1–7]. The network of physically interacting residues within proteins plays two vital roles: they stabilize the protein structure by making numerous short-range van der Waals (vdW) interactions and, importantly, serve as conduits for the long-range transmission of signals. These signals generally involve information transfer or propagation of energy from the binding site of one ligand (protein, DNA or cofactor) to a distal site where another ligand binds or conveys the information of cellular status through post-translational modifications. These so-called perturbations at one site can have positive, negative or no effects (i.e. neutral) at a distal region [8], thus forming the basis of allostery both in the presence or absence of a conformational change [9–11]. While the precise mechanism of information transfer can be varied and is still an open question, it is well recognized that the energetic coupling of residues within a protein—in other words, the intra-protein interaction network—plays a dominant role in this process [1,12–15].

A well-established and heuristic method to infer energetic coupling is the statistical coupling analysis (SCA) [1]. SCA involves a combination of multiple sequence alignment (MSA) of a particular protein family and a perturbative analysis of the MSA to infer coupled residues through sequence

Received: 20 April 2017
Revised: 12 May 2017
Accepted: 17 May 2017Accepted Manuscript online:
18 May 2017
Version of Record published:
6 July 2017

co-variation. Other avenues involve experimentally intensive double-mutant cycles [12] or a combination of elastic network models of proteins and normal mode analysis [16–18]. While these methods have contributed to the understanding of specific systems, an inclusive answer to the question on the extent and magnitude of energetic or thermodynamic coupling has remained elusive. In particular, modeling the extent of energetic coupling requires an understanding of the strength and distribution of the local and non-local interactions within a protein structure, and this is explicitly available through mutational studies. It is also well known that most mutational effects, particularly those that perturb the interaction network within the protein interior, are destabilizing [19] (perturbation of solvent-exposed charged residues can result in non-trivial effects [20,21]). Reproducing these thermodynamic signatures is therefore a critical requirement for any model as the same set of interactions that are perturbed upon mutations are the ones that ‘transmit’ signals upon ligand or protein binding or post-translational modifications [5].

In this regard, employing a multi-model approach involving interaction network analysis using network theory and all-atom molecular dynamics (MD) simulations, we recently showed that the effect of mutations propagate into the structure as far as ~15–20 Å from the mutated site and dissipate exponentially mimicking experimental observations [22]. Simplifying the distance dependence to two interaction shells around a perturbed residue (see Supplementary Figure S1A, for a simple schematic) and recasting these results into the Wako-Saitô-Muñoz-Eaton (WSME) model [23,24], it was possible to estimate the partitioning of destabilization energy in the first (x_1) and second shell (x_2) of interactions (radius 6 Å each) [22]. The partitioning is found to be 0.50:0.20 ($x_1:x_2$) and is also weighted by the nature of the mutation (Mut). These observations can be concisely represented in the form of an empirical relation below with respect to the perturbed wild-type (WT) residue i :

$$\begin{aligned}\Delta Q_{i,j} &= Q_{i,j}^{\text{WT}} x_1 (1 - n_{\text{Mut}}/n_{\text{WT}}) \\ \Delta Q_{j,k} &= Q_{j,k}^{\text{WT}} x_2 (1 - n_{\text{Mut}}/n_{\text{WT}})\end{aligned}\tag{1}$$

where $Q_{a,b}$ is the number of heavy atom contacts within a specific cut-off (say, 6 Å) between residues a and b that is obtained from a contact map; ΔQ represents the loss of interactions (or the derived perturbations) between pairs of interactions between neighbors in the first shell of residue i (defined by j), and between the first shell and second shell neighbors (k); and n is the number of atoms corresponding to residue i in the wild-type (n_{WT}) or in the mutated residue (n_{Mut}), respectively (Figure 1A). An equivalent version of equation (1) in terms of van der Waals interaction energy is provided in reference [22]. This structural perturbation approach is able to reproduce the changes in stabilities induced by 375 point-mutations in 19 different protein structures, highlighting that it is faithful to protein destabilization thermodynamics [22]. Importantly, it suggests that the effect of any perturbation percolates into the second shell of interactions and also dissipates as $x_2 < x_1$.

Because the interaction network serves as a conduit for signals, the empirical approach discussed above (eq. 1) can potentially shed light on the degree of coupling between protein residues. Does it therefore capture the exponential dissipation of mutational effects observed in network analysis, MD simulations and experiments? [22] Does the observed propagation obtained from the perturbation of residues in just a single structure explain the allosteric coupling between residues gleaned from the SCA that requires hundreds of sequences? Understanding the extent of energetic coupling would also shed light on the remarkably cooperative folding behaviors of proteins that has its origins in the nature of the interaction network. In fact, it is well established from theory [25], off-lattice models of proteins [26,27] and detailed NMR experiments [28,29] that packing effects arising from both the first shell of interactions around a residue or even the second shell (i.e. neighbors of first shell residues) contribute to the cooperative folding of proteins. Signaling through the interaction network within proteins and folding cooperativity are therefore two distinct but related aspects of the same problem, and understanding one will necessarily reveal insights into the other. Specifically, can the perturbation approach (eq. 1) predict the degree of residue-level thermodynamic coupling gleaned from detailed NMR experiments that employ temperature as a perturbation?

Methods

Predicting the changes in thermodynamic stability induced by mutations

The changes in protein stability upon point mutations in the hydrophobic core of proteins are predicted using the Ising-like Wako-Saitô-Muñoz-Eaton (WSME) model [23,24] with energy terms for packing, solvation and

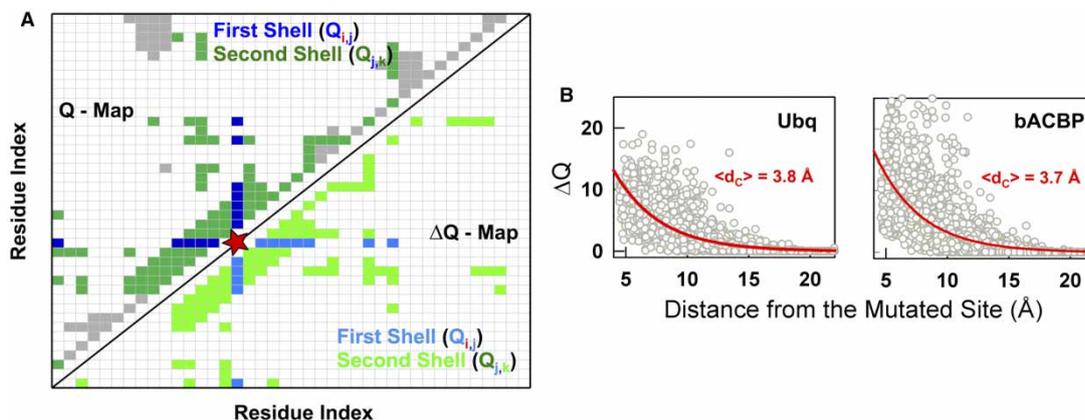


Figure 1. The structural perturbation approach reveals an exponential dissipation of mutational effects.

(A) A schematic of the contact map (Q map; upper triangular matrix) and the ΔQ map (lower triangular matrix). A uniform-color code for the first- and second-shell is employed for the sake of clarity. The mutated residue is shown in red while the first- and second-shell neighbors are shown in dark blue and dark green, respectively, in the upper triangular matrix (also see Supplementary Figure S1A for a simplified schematic). Gray squares represent interactions that are unperturbed. Upon introducing a perturbation according to equation (1), the number of contacts of only the first- and second-shell neighbors of the perturbed residue is decreased by a factor, resulting in the ΔQ map (lower triangular matrix). (B) A plot of the propagation of perturbations as seen by ΔQ as a function of distance from the perturbed site for every single residue of ubiquitin (PDB id: 1UBQ) and bACBP (2ABD). $C_\alpha - C_\alpha$ distances are employed here and throughout the manuscript. The large scatter in panel B is a result of the diversity in the residue environment that is explained in Figure 2. The exponential fits are, however, well defined with a fitting error of ~ 0.06 Å.

electrostatics, following the protocol prescribed previously [20,30]. The following four parameters are fixed for all the proteins: van der Waals interaction (vdW) energy per native heavy-atom contact from the contact map ($\xi = 70.5 \text{ J mol}^{-1}$); heat capacity change per native contact ($\Delta C_p^{\text{cont}} = -0.36 \text{ J mol}^{-1} \text{ K}^{-1}$); entropic cost for fixing a residue in the native conformation ($\Delta S_{\text{conf}} = -16.5 \text{ J mol}^{-1} \text{ K}^{-1}$ per residue); and the m -value per residue for deriving a chemical unfolding curve ($m_{\text{res}} = 0.1 \text{ kJ mol}^{-1} \text{ M}^{-1}$), apart from the pH (7.0) and ionic strength conditions (0.05 M). Since we are interested in predicting only the relative changes in stability upon point mutations, the absolute values of the parameters are not critical. Here, we employ the average numbers from a previous detailed analysis [22]. An all heavy-atom based contact map with a distance cut-off of 6 Å, that excludes nearest neighbors is used for calculating van der Waals interactions. This contact map is provided as an input to calculate total partition function employing the transfer-matrix formalism of Wako and Saitô [23] from which an unfolding curve is derived. The unfolding curves thus generated are then fitted to a two-state model, thus yielding the stability at zero denaturant concentration for the wild-type protein (ΔG_{WT}). A more detailed description of the model and its energy terms can be obtained from our previous works [20,30,31].

The mutational effects are predicted (i.e. ΔG_{mut}) by using identical parameters to those above and by modulating the van der Waals interactions of the mutant using the fixed relation given below:

$$\begin{aligned}
 E_{\text{vdW}}^{\text{mut}(i,j)} &= E_{\text{vdW}}^{\text{WT}(i,j)} (1 - x_1 (1 - n_{\text{mut}}/n_{\text{WT}})) \\
 E_{\text{vdW}}^{\text{mut}(j,k)} &= E_{\text{vdW}}^{\text{WT}(j,k)} (1 - x_2 (1 - n_{\text{mut}}/n_{\text{WT}}))
 \end{aligned}
 \tag{2}$$

where i is the mutated residue, j refers to first shell neighbors and k refers to second shell neighbors of the mutated residue. The pair-wise van der Waals interaction energies (E) are weighted by the nature of the mutation (the number of atoms in mutated residue n_{mut} or wild type n_{WT}) and an empirical factor (x_1 or x_2) to denote the extent of fractional destabilization ($0 < x_1, x_2 < 1$) in the first and second shells (Supplementary Figure S1A). A value of 0.5 and 0.2 was employed for x_1 and x_2 , respectively, and with equal shell radii of 6 Å as prescribed previously [22]. This was arrived at from a detailed analysis of more than 20,625 unfolding curves and different first-/second-shell radii (including 6 and 6 Å, 6 and 5 Å, 6 and 4 Å, 5 and 5 Å, 5 and 4 Å

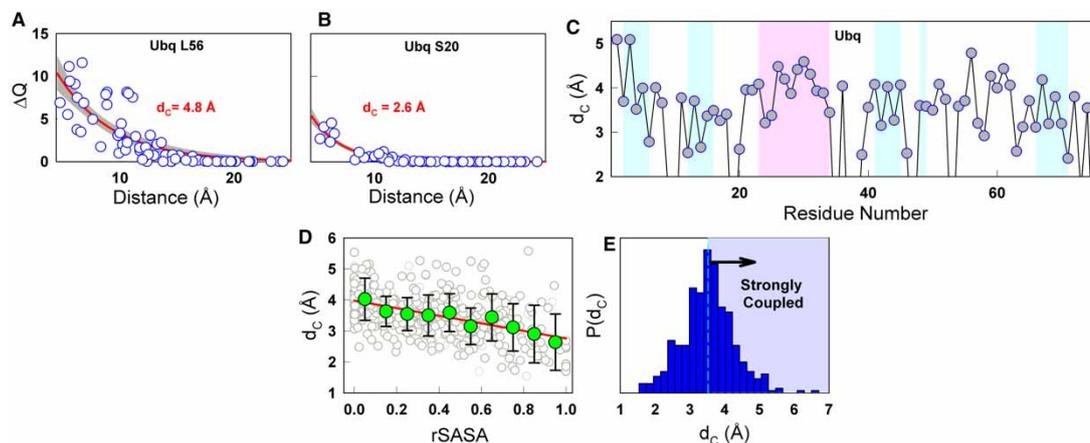


Figure 2. Understanding the coupling distance.

(A and B) Representative examples of the differences in the exponential dissipation upon perturbation of a buried (panel A) and a solvent-exposed residue (panel B). The exponential fit is shown in red together with the 95% confidence interval (shaded areas). (C) Residue-wise coupling distances of ubiquitin. The cyan and pink shades represent beta-strands and alpha helices, respectively. (D) The dependence of coupling distance on the relative solvent accessible surface area (rSASA) extracted from a perturbation analysis of six proteins used in this study. The red line is a linear fit to the data in gray circles while the green circles represent the averages over rSASA bins of size 0.1 units. (E) Distribution of coupling distances. Residues with d_c higher than the mean of 3.5 Å (shaded area) are potentially strongly coupled to their neighbors. It is important to note that only the empirical relations in equation (1), apart from the contact map, are required for predicting the d_c values.

apart from various x_1 and x_2 combinations), which was capable of reproducing the mutational destabilizations involving 375 mutations from 19 proteins [22].

Rapid alanine scanning mutagenesis

Equation (2) can also be recast in terms of the pair-wise contact map (Q) and is provided in the main text (eq. 1). This essentially liberates the empirical relation from the WSME model and allows a scan of the environment of every residue in a rapid manner by inputting just the contact map and the identity of the mutation. Specifically, for an alanine scanning mutagenesis, we mutate every residue to alanine (and in the case of alanine, to glycine), except for proline and glycine, on the six proteins studied (see Supplementary Figure S1B for a flowchart). The effective change in the number of interactions expected for a given residue, ΔQ , is estimated by summing up the pair-wise terms in equation (1) (equivalent to the changes in the van der Waals interaction energy between the wild-type and mutant). This is now plotted as a function of distance from the mutated site for every residue. Note that ΔQ is always positive in this method simply because we truncate the residue to alanine as is always done in experimental alanine-scanning mutagenesis. Potential changes to the structure introduced through mutations or local stabilizations will not be captured as the method relies solely on the contact map of the WT.

Results

Understanding the coupling distance

We first perform an *in silico* alanine scanning mutagenesis following the empirical relation outlined in equation (1) on six proteins with varied secondary structural contents and topologies (bACBP, Im9, SH3, Ubq, FKBP12 and Sso7d). We find that the decay of ΔQ is exponential-like from the mutated site with a mean distance constant or a coupling distance (d_c) of 3.8 Å and 3.7 for Ubq and bACBP, respectively (Figure 1B and Supplementary Figure S2). The magnitude of d_c is of the same order as observed from the interaction network perturbation analysis and MD simulations of several mutants of ubiquitin [22] ($d_c \sim 4.1$ – 4.7 Å), despite this dependence not being explicitly included in the perturbation equation. Moreover, the mean d_c for each of the proteins is also $\sim 3.8 \pm 0.06$ Å indicating that the average extent of coupling is largely independent of protein size or structural class (Supplementary Figure S2). It is also important to note that this structural perturbation

is consistent with thermodynamic experiments, as we obtain an average correlation of ~ 0.8 between the experimental and predicted destabilization ($\Delta\Delta G$) for the six proteins when equation (1) is cast into the WSME model (Supplementary Figure S3).

The average d_C can be de-convoluted into contributions from individual residues that range from ~ 2 Å to ~ 5 Å for all proteins contributing to the apparent large scatter in Figure 1B (Figure 2A,B). The extent of percolation into the protein structure (i.e. the distance at which no effect is felt) in these cases corresponds to ~ 8 and ~ 17 Å, respectively. Figure 2A plots the ΔQ vs. distance plot for a buried residue L56 in ubiquitin that displays a coupling distance of ~ 4.8 Å that is much higher than the average. A solvent-exposed residue S20, on the other hand, exhibits a smaller coupling distance (Figure 2B). These observations are intuitively expected, as residues whose side-chains are pointing within the hydrophobic core will be interacting with an array of residues in the crowded protein interior. The situation is the exact opposite for a solvent-exposed residue that is merely coupled to residues on the surface. In other words, buried residues will contribute to large destabilization and hence will be more conserved over evolutionary time (to avoid potential detrimental effects) compared with solvent-exposed residues (where the mutational effect is small). This is exactly what is observed in sequence-structure-based studies of protein evolution that point to a large conservation of buried residues and a weaker conservation of solvent-exposed residues [32]. This enables a rapid exploration of sequence-function space during evolution without significantly affecting the protein stability.

A plot of the magnitude of d_C as a function of sequence index reveals no apparent dependence on the protein class or topology (Figure 2C and Supplementary Figure S4). However, we find a weak correlation ($r \sim 0.56$ and explains just $\sim 31\%$ of the variance in the dataset) between relative solvent accessible surface area (rSASA) of residues in the protein structure and the corresponding coupling distances. This holds true irrespective of a global correlation or binning analysis (green in Figure 2D). This is also evident in Figure 2D where we find a large spread in the coupling distances even at low rSASA, indicating that d_C is not just a function of rSASA. One reason for only a weak correlation is that the coupling distance is very sensitive to the local packing environment and the size of the amino acid. For example, alanine will be coupled to only a few distant residues even if it is completely buried, since the side-chain is short. The opposite holds true for a large residue that is exposed to the solvent. It can be seen that different secondary-structure elements by themselves exhibit different mean coupling magnitudes. Even within a single secondary structure, a unique pattern is evident where the d_C magnitude fluctuates between high and low; for example, adjacent residues in beta strands point to opposite coupling magnitudes while the helices display a similar pattern but with a weaker dependence (Supplementary Figure S5). These differences have their origins in the organization of residues within the corresponding secondary-structure elements: adjacent residues in beta strands have their side-chains pointing in opposite directions and if the residue is buried then it exhibits larger coupling and vice versa (Supplementary Figure S5). In other words, the coupling distance is a robust property of protein residues—since it is sensitive to the local packing environment, secondary structure, the nature of the amino acid and the relative solvent accessibility—and can potentially highlight strongly coupled residues (Figure 2E). Note that ‘strong coupling’ is a relative term and it will depend on the immediate environment of the different residues in the protein.

Comparison with SCA predictions

The structural perturbation method we present here shows that mutating a residue at a particular site can have non-intuitive effects since most residues are coupled to residues even in their second shell, i.e. higher order effects. It therefore serves as a very powerful and a rapid avenue to probe for the extent of energetic or allosteric coupling in proteins that has been proposed to be determine dynamic allostery [11] (responses at distal sites in the absence of conformational change). It should therefore be possible to recapitulate SCA observations by a mere perturbation of specific residues that can also be functionally important.

The SCA method was first applied to the protein PDZ3 (Figure 3A), which plays a vital role in cell signaling by aiding the assembly of multi-protein complexes [1]. We compare the results of our perturbation approach to the more complex SCA. We perturb just four ligand-binding residues (F325, N326, H372, K380) and obtain a similar exponential dissipation of their effects with a mean coupling distance of 4.3 Å (blue in Figure 3B). Remarkably, this approach in itself captures most of the residues identified as energetically coupled from the SCA [1] (cyan circles in Figure 3B). The same holds true for the enzyme DHFR (dihydrofolate reductase; Figure 3C), which binds to the substrate dihydrofolate and reduces it to tetrahydrofolate aided by the cofactors NADPH and NADP⁺. SCA of DHFR has shown nearly 50 residues in the structure to directly affect enzyme activity [33]. To see if these observations can be reproduced from a simple distance dependence of the

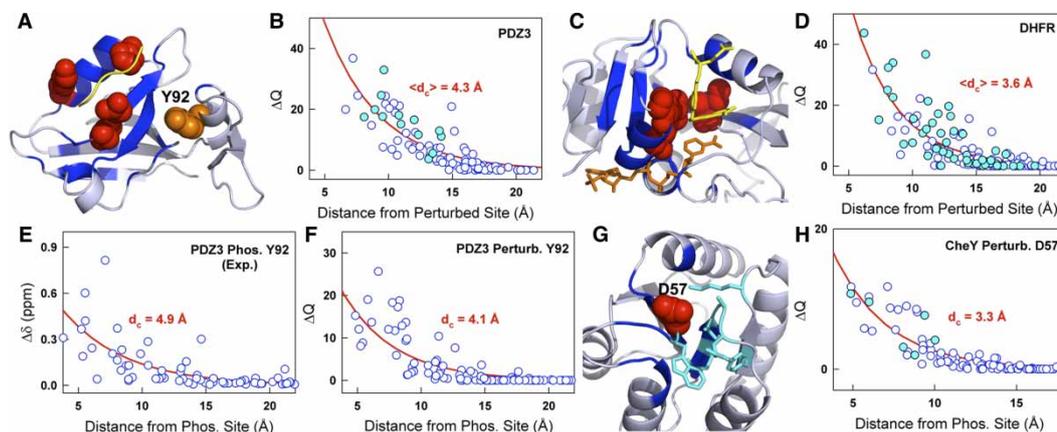


Figure 3. Identifying allosterically coupled residues via perturbations.

Red curves in panels **B**, **D**, **E**, **F** and **H** represent exponential fits. **(A)** Cartoon of PDZ3 (PDB id: 1BFE) with the perturbed residues around the bound peptide (yellow) in red and the residues identified as energetically coupled (by employing eq. 1) in blue. **(B)** The effective distance dependence of ΔQ upon perturbing the residues shown in red in panel **A**. Coupled residues as identified by the Statistical Coupling Analysis (SCA) are shown in cyan. **(C)** Structure of DHFR (1RX2) with the perturbed residues in red, coupled residues in blue, the substrate folate in yellow and the co-factor NADP in orange. **(D)** Same as panel **B**, but for DHFR. **(E)** Difference in backbone chemical shifts between the phosphorylated and non-phosphorylated PDZ3 as a function of distance from the C_{α} of Y92. **(F)** Distance dependence extracted from the perturbation of Y92. **(G)** Model of CheY (1F4V) with the residues coupled to D57 (red) in blue and cyan. The allosteric quartet of residues is shown as sticks. **(H)** Same as panel **F** but for CheY.

dissipation of perturbation, we perturb four spatially contacting residues close to the ligands (I5, T46, I50 and I94, of which only T46 is highly conserved) resulting in a d_C of 3.6 Å (Figure 3D). This perturbation approach again captures nearly 70% of the residues identified as being energetically coupled from the SCA (cyan circles in Figure 3D).

Rationalizing the allosteric modulations of PTMs

A related phenomenon is observed in a post-translational modification (PTM) event that takes place in PDZ3. Phosphorylation of Y92 in PDZ3 has been shown to allosterically modulate the binding affinity of a peptide ligand at a distant site [34]. Since a detailed chemical shift mapping is available from experiments, we plot the difference in chemical shifts of backbone atoms between the WT and the phosphorylated variant as a function of distance from Y92. We find that the backbone chemical shifts of distant residues are also modulated, effectively resulting in an exponential dependence with a coupling distance of 4.9 Å (Figure 3E). Perturbation analysis of PDZ3 Y92 (without considering charge-induced structural changes) reveals a d_C of 4.1 Å (Figure 3F). While the experimental backbone chemical shifts cannot be directly compared to the predicted perturbations (ΔQ), the extent of propagation is very similar, suggesting that the experimentally observed allosteric modulation is a result of the propagation and dissipation of the phosphorylation event into the structure. CheY, which plays a vital role in bacterial chemotaxis, also undergoes a phosphorylation event at D57 that in turn determines the direction of flagellar rotation through regulation of binding (Figure 3G) [35]. A perturbation analysis of D57 results in a low coupling distance of 3.3 Å (since it is solvent exposed; Figure 3H), but the sphere of influence of perturbation still encompasses the known allosteric quartet of residues (including Y106, whose rotameric status is modified; residues in cyan sticks in Figure 3G).

Insights into protein folding cooperativity

NMR experiments that monitor changes in the chemical environment of individual atoms in proteins with temperature indicate that higher-order couplings emerge through interactions mediated by the first-shell contacts [28,29]. Since our method explicitly accounts for these effects, we map the changes in the number of contacts (ΔQ) upon perturbation of every residue in the form of a global ΔQ map (bottom right triangle of

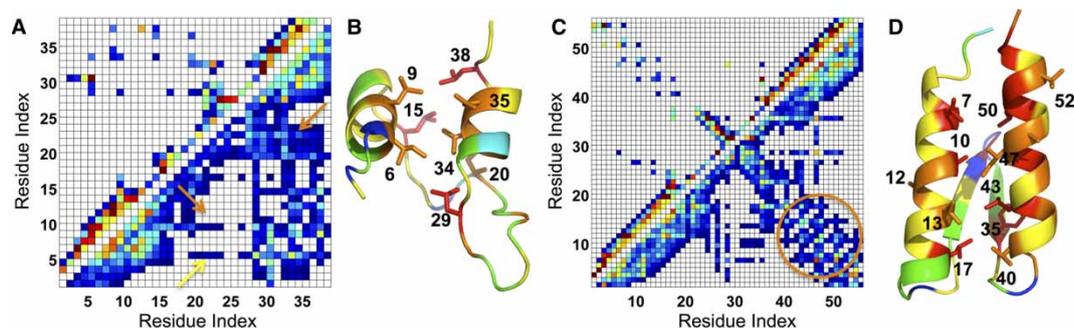


Figure 4. Higher-order interactions from ΔQ map.

The spectral color-coding in all panels goes from blue to red representing weak to strong interactions or energetic coupling (blue to red). **(A)** The contact map (upper left triangle) and the ΔQ map of BBL (bottom right triangle). The orange arrows represent experimentally observed long-range couplings while the yellow arrow represents the couplings observed only in the perturbation approach. **(B)** The d_C -map of BBL (2CYU) with the strongly coupled residues numbered. **(C and D)** Same as the first two panels but for gpW (1HYW). Note the alternating patterns in the cartoons originating from a combination of residue size and the surrounding packing environment.

Figure 4A,C). This is identical to the approach taken for the lower triangular matrix of **Figure 1**; in this case, however, the ΔQ distribution from every residue is summed up to provide a global ΔQ map. We observe a pattern very similar to the interpretation of NMR experiments wherein additional spots expectedly appear on the ΔQ map for both BBL and gpW (**Figure 4A,C**). The ΔQ map is more discrete compared with the continuous pattern of experimental coupling indices, as it is derived from a discrete contact map (upper left triangles in **Figures 4A** and **4C**). Residues with large coupling distances are concentrated in the hydrophobic core, as noted before (**Figure 4B**). For example, in BBL, we observe that L9 and several residues of the second helix mediate an array of non-local contacts exactly as seen in experiments [28] (orange arrows in **Figure 4A**). Interestingly, I6 seems to interact strongly with several residues farther along the sequence (yellow arrow in **Figure 4A**); however, this is not observed in experiments. This possibly arises from the use of a single structure as a representative of possibly a heterogeneous native ensemble, thus over-estimating the coupling.

In gpW, we observe several long-range couplings, particularly between the two strands and the rest of the structure, which are similar to experimental observations [29] (orange circle in **Figure 4C**). However, the two strands of gpW exhibit low coupling ($\langle d_C \rangle$ of ~ 3.5 Å), since they interact only weakly with the hydrophobic core defined by the two helices. This weak coupling potentially contributes to the population of a partially structured state during the folding of gpW in which both the strands are disordered [36]. The second helix has a higher mean coupling distance (more oranges and reds in **Figure 4D**; $\langle d_C \rangle \sim 4.1$ Å) compared with the first helix ($\langle d_C \rangle \sim 3.8$ Å) consistent with NMR experiments that indicate the second helix to be more stable. These results highlight that variations in energetic coupling within a single protein domain determine the effective folding cooperativity, as previously noted. Furthermore, the large dispersion in the ΔQ map suggests that a residue can be considered as ‘fully folded’ only when the second-shell of interactions is also formed; this aspect is generally not considered in protein folding models, including the ones employed by us (i.e. the WSME model).

Discussion

The residue-level structural perturbation approach we propose here simultaneously explains experimental protein destabilization energetics and the results of sequence-based statistical coupling analysis. This self-consistency, we believe, is an important requirement to model dynamic allostery since signal transmission necessarily takes place through the network of interactions within proteins. The propagative effect is not just restricted to mutations, but is also surprisingly consistent with the effects of ligand binding and post-translational modifications on proteins. Our results further explain a recent meta-analysis of the conservation patterns in enzymes [37]; it was found that residues up until 20–25 Å away from the active site exhibit minimal mutational rate compared with the rest of the structure, suggestive of long-range coupling among residues. We find that such couplings arise merely from a combination of distance proximity, as originally proposed [12],

and the propagative effect of perturbations in an interaction network. This could therefore be a generic feature of proteins or systems with multiple weak non-covalent interactions. Our observations also highlight that there are potentially numerous allosteric sites in enzymes since any residue within ~ 15 Å from the active site is expected to influence enzyme activity. Our method therefore opens the door for rapidly identifying specific solvent-exposed sites or ‘hot spots’ that a drug molecule can bind to in order to regulate enzyme function without directly binding to the active site.

Both the SCA and the method proposed here are computational approaches and it is therefore unreasonable to expect a perfect correlation between the two. However, the large overlap between the predictions of the SCA (from numerous sequences) and that gleaned from a simple structural perturbation (Figure 3) suggests that our perturbation approach can be employed as a first step to identify potentially coupled residues that can further be refined by either SCA or more directly through experiments. In elastic network models (ENMs), it is generally assumed that residues separated by up to 10–15 Å are coupled through springs of either a fixed spring constant or that decay with increasing distance. Interestingly, incorporation of a distance-dependent spring constant provides a better agreement with experiments [38]. The results of our perturbation approach points to a simple reason as to why such an assumption provides a better agreement, since the effect of any perturbation is expected to decay the farther a particular residue is from a perturbed site (the coupling is weaker; also see ref. [22]). It is also important to note that the perturbation through alanine scanning is purely employed as a tool to probe the underlying features of the intra-molecular interaction network; the resulting output is ΔQ as a function of residue index (or distance from the perturbed residue), which is a measure of the extent to which the perturbed residue is coupled to its neighbors, i.e. its immediate environment. Our method therefore does not report on possible local stabilization (negative ΔQ) of the protein structure. Moreover, since the perturbation approach relies on the contact map of the WT as the input, any potential changes to the structure upon ligand binding or mutation (e.g., small to large side-chain substitutions) will not be accounted for.

We also highlight how cooperativity is built into protein systems through weak higher order interactions by explicitly decoupling the environment of a residue into two shells of interactions and destabilization energies. Many-body terms (or the second-shell effects) in protein folding models should therefore provide a more realistic description of the (un)folding processes. This however, requires the folding community to consistently move beyond contact maps and introduce residue-environment-specific energy terms in structure-based models, similar to the amino acid ‘burial’ code pioneered by de Araujo and Onuchic [39]. However, a more detailed understanding of energetic coupling and the associated timescales from multiple systems is required to paint a quantitative picture on the modulations of binding affinity at a distant site (the ‘holy grail’ of allostery) or to predict the degree of folding cooperativity from first principles. The experimentally derived thermodynamically consistent approach we present here that recasts the residue environment into a single number provides a stepping-stone in that direction.

Abbreviations

MD, Molecular dynamics; NMR, Nuclear magnetic resonance; PTM, Post-translational modifications; SCA, statistical coupling analysis; WSME, Wako-Saitô-Muñoz-Eaton.

Author contribution

N. R and A. N. N. designed the approach, analyzed the data and wrote the paper.

Acknowledgements

A. N. N. is a Wellcome Trust/ DBT India Alliance Intermediate Fellow.

Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

References

- 1 Lockless, S.W. and Ranganathan, R. (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295–299 doi:10.1126/science.286.5438.295
- 2 Gunasekaran, K., Ma, B. and Nussinov, R. (2004) Is allostery an intrinsic property of all dynamic proteins? *Proteins* **57**, 433–443 doi:10.1002/prot.20232
- 3 Fuentes, E.J., Der, C.J. and Lee, A.L. (2004) Ligand-dependent dynamics and intramolecular signaling in a PDZ domain. *J. Mol. Biol.* **335**, 1105–1115 doi:10.1016/j.jmb.2003.11.010

- 4 Whitten, S.T., Garcia-Moreno E., B. and Hilser, V.J. (2005) Local conformational fluctuations can modulate the coupling between proton binding and global structural transitions in proteins. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4282–4287 doi:10.1073/pnas.0407499102
- 5 Liu, T., Whitten, S.T. and Hilser, V.J. (2006) Ensemble-based signatures of energy propagation in proteins: a new view of an old phenomenon. *Proteins* **62**, 728–738 doi:10.1002/prot.20749
- 6 Tzeng, S.-R. and Kalodimos, C.G. (2011) Protein dynamics and allostery: an NMR view. *Curr. Opin. Struct. Biol.* **21**, 62–67 doi:10.1016/j.sbi.2010.10.007
- 7 McLaughlin, Jr, R.N., Poelwijk, F.J., Raman, A., Gosal, W.S. and Ranganathan, R. (2012) The spatial architecture of protein function and adaptation. *Nature* **491**, 138–142 doi:10.1038/nature11500
- 8 Motlagh, H.N. and Hilser, V.J. (2012) Agonism/antagonism switching in allosteric ensembles. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 4134–4139 doi:10.1073/pnas.1120519109
- 9 Monod, J., Wyman, J. and Changeux, J.-P. (1965) On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12**, 88–118 doi:10.1016/S0022-2836(65)80285-6
- 10 Koshland, Jr, D.E., Némethy, G. and Filmer, D. (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **5**, 365–385 doi:10.1021/bi00865a047
- 11 Cooper, A. and Dryden, D.T.F. (1984) Allostery without conformational change: a plausible model. *Eur. Biophys. J.* **11**, 103–109 doi:10.1007/BF00276625
- 12 Fodor, A.A. and Aldrich, R.W. (2004) On evolutionary conservation of thermodynamic coupling in proteins. *J. Biol. Chem.* **279**, 19046–19050 doi:10.1074/jbc.M402560200
- 13 Liu, Z., Chen, J. and Thirumalai, D. (2009) On the accuracy of inferring energetic coupling between distant sites in protein families from evolutionary imprints: illustrations using lattice model. *Proteins* **77**, 823–831 doi:10.1002/prot.22498
- 14 DuBay, K.H., Bowman, G.R. and Geissler, P.L. (2015) Fluctuations within folded proteins: implications for thermodynamic and allosteric regulation. *Acc. Chem. Res.* **48**, 1098–1105 doi:10.1021/ar500351b
- 15 Doshi, U., Holliday, M.J., Eisenmesser, E.Z. and Hamelberg, D. (2016) Dynamical network of residue–residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4735–4740 doi:10.1073/pnas.1523573113
- 16 Bahar, I., Erman, B., Haliloglu, T. and Jernigan, R.L. (1997) Efficient characterization of collective motions and interresidue correlations in proteins by low-resolution simulations. *Biochemistry* **36**, 13512–13523 doi:10.1021/bi971611f
- 17 Zheng, W., Brooks, B.R. and Thirumalai, D. (2006) Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7664–7669 doi:10.1073/pnas.0510426103
- 18 Haliloglu, T. and Bahar, I. (2015) Adaptability of protein structures to enable functional interactions and evolutionary implications. *Curr. Opin. Struct. Biol.* **35**, 17–23 doi:10.1016/j.sbi.2015.07.007
- 19 Naganathan, A.N. and Muñoz, V. (2010) Insights into protein folding mechanisms from large scale analysis of mutational effects. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8611–8616 doi:10.1073/pnas.1000988107
- 20 Naganathan, A.N. (2012) Predictions from an Ising-like statistical mechanical model on the dynamic and thermodynamic effects of protein surface electrostatics. *J. Chem. Theory Comput.* **8**, 4646–4656 doi:10.1021/ct300676w
- 21 Xiao, S., Patsalo, V., Shan, B., Bi, Y., Green, D.F. and Raleigh, D.P. (2013) Rational modification of protein stability by targeting surface sites leads to complicated results. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11337–11342 doi:10.1073/pnas.1222245110
- 22 Rajasekaran, N., Suresh, S., Gopi, S., Raman, K. and Naganathan, A.N. (2017) A general mechanism for the propagation of mutational effects in proteins. *Biochemistry* **56**, 294–305 doi:10.1021/acs.biochem.6b00798
- 23 Wako, H. and Saitō, N. (1978) Statistical mechanical theory of the protein conformation. II. Folding pathway for protein. *J. Phys. Soc. Japan* **44**, 1939–1945 doi:10.1143/JPSJ.44.1939
- 24 Muñoz, V. and Eaton, W.A. (1999) A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11311–11316 doi:10.1073/pnas.96.20.11311
- 25 Plotkin, S.S., Wang, J. and Wolynes, P.G. (1997) Statistical mechanics of a correlated energy landscape model for protein folding funnels. *J. Chem. Phys.* **106**, 2932–2948 doi:10.1063/1.473355
- 26 Hilser, V.J., Dowdy, D., Oas, T.G. and Freire, E. (1998) The structural distribution of cooperative interactions in proteins: analysis of the native state ensemble. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9903–9908 doi:10.1073/pnas.95.17.9903
- 27 Shimizu, S. and Chan, H.S. (2002) Anti-cooperativity and cooperativity in hydrophobic interactions: three-body free energy landscapes and comparison with implicit-solvent potential functions for proteins. *Proteins* **48**, 15–30 doi:10.1002/prot.10108
- 28 Sadqi, M., Fushman, D. and Muñoz, V. (2006) Atom-by-atom analysis of global downhill protein folding. *Nature* **442**, 317–321 doi:10.1038/nature04859
- 29 Sborgi, L., Verma, A., Piana, S., Lindorff-Larsen, K., Cerminara, M., Santiveri, C.M. et al. (2015) Interaction networks in protein folding via atomic-resolution experiments and long-time-scale molecular dynamics simulations. *J. Am. Chem. Soc.* **137**, 6506–6516 doi:10.1021/jacs.5b02324
- 30 Naganathan, A.N. (2013) A rapid, ensemble and free energy based method for engineering protein stabilities. *J. Phys. Chem. B* **117**, 4956–4964 doi:10.1021/jp401588x
- 31 Rajasekaran, N., Gopi, S., Narayan, A. and Naganathan, A.N. (2016) Quantifying protein disorder through measures of excess conformational entropy. *J. Phys. Chem. B* **120**, 4341–4350 doi:10.1021/acs.jpcc.6b00658
- 32 Lin, Y.-S., Hsu, W.-L., Hwang, J.-K. and Li, W.-H. (2007) Proportion of solvent-exposed amino acids in a protein and rate of protein evolution. *Mol. Biol. Evol.* **24**, 1005–1011 doi:10.1093/molbev/msm019
- 33 Reynolds, K.A., McLaughlin, R.N. and Ranganathan, R. (2011) Hot spots for allosteric regulation on protein surfaces. *Cell* **147**, 1564–1575 doi:10.1016/j.cell.2011.10.049
- 34 Zhang, J., Petit, C.M., King, D.S. and Lee, A.L. (2011) Phosphorylation of a PDZ domain extension modulates binding affinity and interdomain interactions in postsynaptic density-95 (PSD-95) protein, a membrane-associated guanylate kinase (MAGUK). *J. Biol. Chem.* **286**, 41776–41785 doi:10.1074/jbc.M111.272583

- 35 McDonald, L.R., Whitley, M.J., Boyer, J.A. and Lee, A.L. (2013) Colocalization of fast and slow timescale dynamics in the allosteric signaling protein CheY. *J. Mol. Biol.* **425**, 2372–2381 doi:10.1016/j.jmb.2013.04.029
- 36 Sanchez-Medina, C., Sekhar, A., Vallurupalli, P., Cerminara, M., Muñoz, V. and Kay, L.E. (2014) Probing the free energy landscape of the fast-folding gpW protein by relaxation dispersion NMR. *J. Am. Chem. Soc.* **136**, 7444–7451 doi:10.1021/ja502705y
- 37 Jack, B.R., Meyer, A.G., Echave, J. and Wilke, C.O. (2016) Functional sites induce long-range evolutionary constraints in enzymes. *PLoS Biol.* **14**, e1002452 doi:10.1371/journal.pbio.1002452
- 38 Riccardi, D., Cui, Q. and Phillips, Jr, G.N. (2009) Application of elastic network models to proteins in the crystalline state. *Biophys. J.* **96**, 464–475 doi:10.1016/j.bpj.2008.10.010
- 39 Pereira de Araujo, A.F. and Onuchic, J.N. (2009) A sequence-compatible amount of native burial information is sufficient for determining the structure of small globular proteins. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19001–19004 doi:10.1073/pnas.0910851106