Ensemble origins and distance-dependence of long-range mutational effects in proteins

Adithi Kannan, Athi N. Naganathan
athi@iitm.ac.in

Highlights

Statistical mechanical modeling predicts long-range mutational effects in proteins

- Coupling free energies follow an exponential trend with distance from mutated site
- Mutations alter the distribution of states that manifests as distance dependence
- The proposed method can be employed to understand dynamic allostery

Kannan & Naganathan, *iScience* 25, 105181
October 21, 2022 © 2022 The Author(s).
https://doi.org/10.1016/j.isci.2022.105181

**Effect:** Charged > Hydrophobic > Hydrophilic mutations
**Mutational effect till 15-20 Å**
**Coupled residues**
**Distance from mutated site**
**Population redistribution**
Ensemble origins and distance-dependence of long-range mutational effects in proteins

Adithi Kannan¹ and Athi N. Naganathan¹,²,*

SUMMARY
Mutational effects in globular proteins exhibit an exponential-like decreasing dependence on distance from the mutated site, suggestive of long-range modulation of structural-thermodynamic features. Here, we extract the physical origins of this pattern by employing a statistical-mechanical model to construct conformational ensembles of three archetypal proteins. Through large-scale in silico alanine-scanning mutagenesis, we show that inter-residue differential coupling free energies, which are characteristic ensemble thermodynamic properties, follow a similar exponential distance dependence with the effects felt until ~15–20 Å from the mutated site. From the perspective of an ensemble-averaged structure, this feature arises via long-range reorganization of the interaction network on mutations which is more significant for charged residues compared to hydrophobic residues. Our work highlights how subtle alterations in the microscopic distribution of states manifest as a macroscopic distance dependence, the physical origins of mutation-induced dynamic allostery, and the necessity to consider the global intra-protein interaction network to understand mutational outcomes.

INTRODUCTION
Folding-function mechanisms in proteins are primarily explored via mutational perturbations. In the large majority of studies, mutations typically involve the substitution of large hydrophobic residues in the protein core with alanine (Fersht et al., 1992; Hecht et al., 2013; Morrison and Weiss, 2001; Tang and Fenton, 2017; Xu et al., 1998). It is generally assumed that such mutations modulate only the nearest neighbor interactions, i.e. in the first shell around the mutated site, and that the native structure is unperturbed or does not change. However, evidence is accumulating on the long-range effects of even single-point mutations from hydrogen-deuterium exchange experiments (Offenbacher et al., 2017; Pacheco-Garcia et al., 2021; Puri et al., 2022; Roche et al., 2013), NMR measures of order parameters and chemical shift perturbations (Bouvignies et al., 2011; Consonni et al., 1999; Harrinria et al., 2008; Whitley et al., 2008), double-mutant-cycle measures of thermodynamic coupling (Chi et al., 2008; Fodor and Aldrich, 2004), anisotropic/elastica-network models and simulations (Chennubhotla and Bahar, 2007; Gerek and Ozkan, 2011; Xiao et al., 2022; Yu et al., 2019), and statistical-mechanical-cum-perturbation analysis of protein structures (Guarrera and Berezovsky, 2019; Liu et al., 2009; Tee et al., 2020). In addition, while charged residue mutations have been successfully exploited to engineer stabilities (Sanchez-Ruiz and Makhatadze, 2001), they can also alter the folding-conformational landscape due to the long-range nature of charge-charge interactions. Thus, it is increasingly being realized that mutations could implicitly contribute to allosteric or long-range effects, and this could have implications in not just folding mechanisms but also in understanding the evolution of proteins and functionality (Buller et al., 2018; Maria-Solano et al., 2018; Markin et al., 2021; Petrovic et al., 2018; Tokuniki and Tawfik, 2009; Yang et al., 2016; Zayner et al., 2013).

Though mutational outcomes are highly context-dependent, it is observed that changes in certain structural-thermodynamic features on mutations in globular proteins follow a near-universal exponential-like trend when plotted as a function of distance from the mutated site (Rajasekaran et al., 2017a, 2017b). In other words, mutations not only influence the first shell but also the second and even the third shell around the mutated site with the perturbation “felt” until 15–20 Å, but decreasing in an exponential manner, at least when monitored from the perspective of NMR experiments (Naganathan, 2019). This observation has been exploited to build an empirical treatment of hydrophobic truncation mutations that simultaneously captures both the two-shell propagation (in the thermodynamic sense) and the
unfolding thermodynamics (stability changes) (Rajasekaran and Naganathan, 2017). These experimental observations and empirical analysis, however, primarily view proteins as a single structure with perturbations “spreading out” radially from the mutation site. In reality, the “native state” is not a single structure but a collection of conformations, the Boltzmann-weighted average of which is observed as NMR or X-ray crystallographic models. The ensemble view of proteins is, in fact, one of the basic tenets of the energy landscape theory of protein folding (Bryngelson et al., 1995), and has been invoked to not only explain protein folding mechanisms but also function, (dynamic) allostery, and epistasis (Biddle et al., 2021; Freire, 1999; Hilser et al., 1998; Luque et al., 2002; Modi et al., 2018; Morrison et al., 2021; Motlagh et al., 2014).

The ensemble view, though attractive, has its associated challenges. How broad are the native ensembles of proteins? Is it possible to define, construct, and refine ensembles on large proteins? Can large-scale in silico mutagenesis be performed to mimic experimental saturation mutagenesis studies? Importantly, can the ensemble view reproduce the exponential-like decay of mutational perturbations when the ensemble effects are mapped on to the native structure? We address these specific questions in the current work by employing the Ising-like statistical mechanical treatment of the folding process termed the Wako-Saito–Muñoz-Eaton (WSME) model (Muñoz and Eaton, 1999; Wako and Saito, 1978). By studying three classic model systems—PDZ (Lockless and Ranganathan, 1999), CheY (Lee, 2015), and CypA (Doshi et al., 2016) (Figure 1A)—and generating more than 300 single-point mutations, we show that subtle alterations in the native ensemble distribution manifest as an exponentially decreasing distance dependence of coupling free energies when mapped on to a single structure. Our work thus reconciles the different viewpoints prevalent in the literature and provides a physical framework for understanding mutational effects and their role in dynamic allostery and protein evolution.

**Figure 1. Extracting coupling free energies to study long-range effects of mutations**

(A) The structure of the three model protein systems. (B) Workflow employed to estimate the positive thermodynamic coupling free energies (coupling FEs) and their analyses. (C–E) The coupling matrices of the wild-type PDZ (panel C), mutant Y397A (panel D), and the Differential Coupling Matrix or DCM (panel E) are depicted. Free energy units are in kJ mol$^{-1}$ with panels C and D plotted with the same color scale (colorbar in panel D). Changes in the coupling FE between distant residues as evident in panel E indicate the allosteric effect of mutations. The extracted positive coupling FE, the free energy profiles, and the free energy of individual substates are compared with respect to the wild type to explore the origins of exponential decay of perturbation with distance.
RESULTS AND DISCUSSION

Differential coupling free energies as proxies for mutational effects

The protocol employed to construct ensembles and quantify the thermodynamic coupling free energies between residues is described in the flowchart presented in Figure 1B (also see “method details” section). Briefly, by employing a binary representation of residue conformational status (1 for folded and 0 for unfolded), specific sequence approximations (single- and double-sequence approximations), and considering sets of two (PDZ and CheY) or three residues (CypA) to fold as one contiguous unit (termed the block, b), the bWSME model (Gopi et al., 2019) constructs an ensemble of $C_{2}^{N_b} + 1 + 2C_{4}^{N_b} + 1$ states where $N_b$ is the total number of blocks spanning the protein sequence. For example, for the 110 residue PDZ, the total number of blocks comes to 50 ($N_b$) when considering a most probable blocksize of 2. This translates to an ensemble of 485,226 conformational substates or microstates—defined by a large array of strings of 1 and 0—from the expression above after excluding specific microstates that are not physically possible (“method details”). The populations are determined by the associated statistical weights of the individual microstates, which are in turn determined by the balance between the favorable (free-)energy terms (van der Waals, electrostatics, and solvation free energy) and destabilizing conformational entropic terms, all calculated from the PDB structure. Accordingly, the total number of microstates for CheY and CypA are 1,011,446 and 721,951, respectively. From the microstate probabilities, partial partition functions are generated for the different number of structured blocks enabling the construction of one-dimensional free energy profiles (1D FEPs). The 1D FEPs enable the description of native ensembles, macroscopic stabilities, and the monitoring of changes in their properties on mutations.

Every wild-type (WT) protein ensemble can be further divided into two sub-ensembles that consider the folded status of one residue with respect to another residue: $\sum p_{i,j}$ accounts for the probabilities of all states in which both residues $i$ and $j$ are folded, while $\sum p_{i,j}$ considers only those states in which residue $i$ is unfolded and $j$ is folded. (Naganathan and Kannan, 2021) Note that this partitioning does not account for states that harbor decoupled residues (Naganathan and Kannan, 2021). Given these, the positive thermodynamic coupling free energy (FE) between two residues can then be calculated from:

$$\Delta G = RT \ln \left( \frac{\sum p_{i,j}}{\sum p_{i,j}} \right)$$

(Equation 1)

By calculating the $\Delta G_i$ for every residue with respect to every other residue, a coupling matrix can be constructed for the WT (Figure 1C) that provides unique information on the regions of the protein structure that are coupled to varying extents in the native ensemble. The coupling FEs have provided extensive insights into the thermodynamic architecture of protein structures as they account for direct interaction between residues (first-shell effects) and beyond the first shell via intermediary interactions and residues (and hence second-shell effects and so on) (Naganathan and Kannan, 2021). Accordingly, we use these ensemble descriptors as the starting point to understand and quantify mutational effects. Alanine-scanning mutagenesis is performed by replacing every residue with alanine (except for proline, glycine, and alanine), resulting in 300 mutations over the three proteins. The mutated structures (mut)—and hence altered contact maps and charge-charge interactions (if the mutation involves a charged residue)—are fed into the model and free-energy profiles and coupling free energies for the mutants ($\Delta G_{\text{mut}}$; Figure 1D) are calculated using identical parameters as the WT. Following this, the differential positive coupling FE matrix, or simply differential coupling matrix (DCM), is generated as $\Delta \Delta G = \Delta G_{\text{mut}} - \Delta G_{\text{WT}}$ (Figure 1E).

Mutational effects are long-ranged and exhibit the characteristic exponential distance-dependence

As representative examples, we generate 1D free-energy profiles (FEPs) for two alanine mutants in each of PDZ, CheY, and CypA (Figure 2). The hydrophobic residue mutations considered are either allosteric hot-spots (Y397 in PDZ, V29 in CypA) or residues that affect the function of their corresponding protein (Y106 in CheY), whereas the charged residues depicted are chosen at random. The FEPs of the mutants are typically destabilized, and this is also observed in the representative mutant set from the lower free energy of the unfolded ensembles (arrows in Figures 2A, 2D, and 2G). Deviations in the folded ensemble on mutations are minimal due to the nature of the collective variable employed to project the conformations (the number of structured blocks), which is dominated by the statistical weight of the state in which the majority of the residues are folded. In some cases, for example in the D160A variant of CypA (red in Figure 2G), the larger
population (or lower free energy) of the partially structured states in the native ensemble are also apparent (red arrow in Figure 2G).

Differential coupling matrices (DCMs; Figures S1–S3) were generated following the protocol described in Figure 1. However, these DCMs cannot be directly compared with experiments. This is because in a typical experiment, one measures a set of residue-level structural-thermodynamic parameters (chemical shifts, typically) for the WT and similarly for the mutant. These are then compared against one another to generate distance-dependent profiles. Seldom are the coupling indices between residues available. To mimic the experimental output, we average over the rows of the DCM to generate residue-level differential coupling indices (DCIs, $C_{ij}^{DD}$) with the WT values as reference. Since we are primarily interested in the overall magnitude of perturbation, only the absolute values of $\Delta \Delta G$ are considered.

The DCIs reveal a pattern wherein the effect of the mutation is the largest around the site of mutation, beyond which it decreases (Figures 2B, 2E, and 2H). The trend can be explained by an exponential function, inspired from experimental observations, of the form $Ae^{-d/d_c} + \delta$ where $A$ is the amplitude of the perturbation (in kJ mol$^{-1}$), $\delta$ is a shift parameter (in kJ mol$^{-1}$), $d$ is the $C_e-C_e$ distance between residues (in Å), and $d_c$ is a characteristic distance of every mutation, which we term as the coupling distance (in Å) (Rajasekaran et al., 2017a). A larger $d_c$ implies that the effect of the mutation is felt at a longer distance from the origin of perturbation. Since $d_c$ is an exponent, the net effect of mutation is negligible only beyond 15–25 Å for $d_c$ values in the range of 5–9 Å. In each of the 6 mutations displayed in Figure 2, the $d_c$ values range between 5.7 and 9.2 with the effect of charged residue mutations “felt” at longer distances compared to mutations involving hydrophobic residues. Mapping the DCIs on to the structure, it is clear that many protein residues are perturbed and to varying extents (Figures 2C, 2F, and 2I).
Exponential distance dependence of mutational effects is independent of the chemical nature of the mutated residue

To derive general principles, we further group the mutations into three classes (hydrophobic, hydrophilic, and charged residues), bin the DCIs, and extract the functional form that best explains the DCIs distance dependence in each of the three proteins. Such grouping ensures sufficient points in every bin, and hence the extracted $d_c$ is more reliable. Mutants involving all the three classes exhibit a similar distance-dependent decay of DCIs that is captured well by a single-exponential function (Figures 3A–3C). In addition, higher order trends begin to emerge with the $d_c$ (charged) > $d_c$ (hydrophobic) > $d_c$ (hydrophilic). This feature is intuitively expected as mutations of charged residues to alanine would modulate both the long-range electrostatics and packing interactions, unlike the other two classes. In the latter classes, since the vast majority of hydrophobic residues are buried within the core of the protein, they exert their effect to longer distances as opposed to the hydrophilic residues that are predominantly solvent-exposed. Given the long-range electrostatics and packing interactions, mutations of charged residues to alanine would modulate both the long-range electrostatics and packing interactions, unlike the other two classes. In the latter classes, since the vast majority of hydrophobic residues are buried within the core of the protein, they exert their effect to longer distances as opposed to the hydrophilic residues that are predominantly solvent-exposed.

![Figure 3. Exponential decay of differential coupling indices irrespective of the class of mutation](image)

The mutated residues are categorized into hydrophobic (HB, blue), hydrophilic (HP, green) and charged residues (CH, red), and their distant dependent decays of the DCIs (in kJ mol$^{-1}$) follow an exponential trend for each of the model proteins (panels A, B, and C). The exponential trend is also evident when the DCIs of a particular class of mutations are grouped, binned and averaged across the three proteins (panels D, E, and F). The shaded region in panels D, E, and F represents the SD obtained upon averaging the DCIs in each bin. The frequency distributions of the coupling distances (panel G) and amplitudes in kJ mol$^{-1}$ (panel H) for each class obtained from the single exponential fit indicate the long-range effect of charged residues followed by hydrophobic and hydrophilic residues. The dashed vertical lines represent the mean value for a particular class of mutations. The absolute values of DCIs averaged across all mutations (panels I, K, and M) and the mapping of these values on the structure (panels J, L, and N) indicate that only certain residues are highly perturbed. The functional sites (depicted in red open circles in panels I and K and yellow circles in panel M), as well as sector residues (depicted in yellow circles in panel I), are less perturbed due to their low coupling with the rest of the structure. The green star symbols represent residues that are both considered as sector and functional site residues of PDZ (panel I), the phosphorylation site of CheY (panel K), and the allosteric site of CypA (panel M). The allosteric quartet residues of CheY are depicted as yellow circles in panel K and sticks in panel L. The functional sites of the three proteins are represented as spheres in panels J, L, and N, whereas the allosteric site is shown in orange spheres.

**Exponential distance dependence of mutational effects is independent of the chemical nature of the mutated residue**

To derive general principles, we further group the mutations into three classes (hydrophobic, hydrophilic, and charged residues), bin the DCIs, and extract the functional form that best explains the DCIs distance dependence in each of the three proteins. Such grouping ensures sufficient points in every bin, and hence the extracted $d_c$ is more reliable. Mutants involving all the three classes exhibit a similar distance-dependent decay of DCIs that is captured well by a single-exponential function (Figures 3A–3C). In addition, higher order trends begin to emerge with the $d_c$ charged $>$ $d_c$ hydrophobic $>$ $d_c$ hydrophilic. This feature is intuitively expected as mutations of charged residues to alanine would modulate both the long-range electrostatics and packing interactions, unlike the other two classes. In the latter classes, since the vast majority of hydrophobic residues are buried within the core of the protein, they exert their effect to longer distances as opposed to the hydrophilic residues that are predominantly solvent-exposed. Given
that the protein-wise trends are internally consistent, a further grouping of the mutational dataset to include only the three classes does not alter the pattern (Figures 3D–3F and S4).

An empirical treatment of the experimentally observed two-shell propagation consistently results in a \( d_L \) of 4–6 Å (Rajasekaran and Naganathan, 2017), unlike experimental trends that point to significantly larger \( d_L \) values (5–12 Å) (Rajasekaran et al., 2017a). To explore if the current unbiased simulations reveal a similar range, we fit the in silico mutation data independently (similar to Figure 2B). The resulting coupling distance ranges are longer (2–10 Å) and follow the trends observed for the grouped data, i.e. mutations involving charged residues exert their effort over longer distances. The amplitudes (Å) follow a similar trend but with the charged residues exhibiting a broader spectrum compared to the other two classes, signaling the possibility of a wider array of effects involving charged-residue mutations. The set of co-evolving residues in PDZ identified via sequence-based analysis has been termed the “sector” residues (Lockless and Ranganathan, 1999). Interestingly, the mean DCIs (averaged across all mutations) at the residue-level indicate a pattern wherein the majority of the weakly coupled residues fall within the sector region (Figures S1 and 3J), highlighting the relation between thermodynamic coupling and function, further recapitulating our original work on the same (Naganathan and Kannan, 2021). A similar observation can also be made for CypA and CheY, with all the active site residues exhibiting a low DCI compared to other regions (Figures 3K–3N).

Mutations reshape the conformational landscape

It is important to emphasize that the observations in Figures 2 and 3 are a manifestation of redistribution of populations in the native conformational landscape. In other words, when a mutation is introduced, it stabilizes some microstates and destabilizes some. This in turn modulates the magnitude of thermodynamic coupling between residues that are already coupled in the native ensemble, the effective average of which is quantified by the DCIs through Equation 1. This can be observed in Figure 4A which plots the difference in free-energy of microstates in the native ensemble for PDZ. The population of a large number of microstates that are partially structured increases on mutations (negative free energy) in Figure 4A calculated as \( \Delta G_{\text{mut}} - \Delta G_{\text{WT}} \) with \( i \) being the index of the microstate or \( \Delta G_{\text{microstate}} \) (Figures 1B and 4), while in some cases, the populations decrease (positive free energy). Similar redistribution of population in the native ensemble is also observed for CheY and CypA (Figure S5). Here, we show only the native ensemble free energy differences to highlight that though the magnitudes are small, they match the differences in free energies predicted by Cooper and Dryden (which are of the order of \(~RT\) or \(~2.5 \text{kJ mol}^{-1}\) required for “dynamic allostery” (Cooper and Dryden, 1984).

Figure 4B plots the distribution of the mean difference in free energy between the WT, i.e. the distribution of \( < \Delta G_{\text{mut}} - \Delta G_{\text{WT}} > \) averaged across all mutations for a particular microstate. The average effect of a mutation is destabilizing as we perform only alanine-scanning mutagenesis. The native ensemble effects highlighted in Figure 4A constitute only a small portion of the overall landscape (Figures 4B and 4C). There are combinatorially more substates possible in the partially structured ensemble, and a large fraction of them are stabilized (compared to the fully folded microstate). The overall effects are very similar for the distributions for CypA and CheY (Figures 4D and 4E). These plots reveal that the large distribution of states modulated on mutations enables access to partially structured states that are otherwise less accessible in the native ensemble. Thus, mutations invariably modulate populations of substates both in the native ensemble and in the partially structured ensemble, potentially contributing to functional tuning. Even in the absence of functional changes, the differential coupling matrix of a mutant is never the same as the WT (as Equation 1 will always be different), particularly in alanine-scanning mutagenesis. We would also like to reiterate that such changes at the level of microstates are never observable in experiments.

To summarize, we construct ensembles of microstates from the WSME model starting from a single native structure, introduce in silico mutations, and quantify the differences in populations via coupling matrices. The resulting ensemble-averaged DCIs display a trend that mimics experimentally observed distance-dependent exponential-like decay of thermodynamic features (Figure 4F). It is apparent that population redistributions in the native landscape can contribute to macroscopic exponentially decaying distance-dependent effects of protein structural-thermodynamic features. A unique quantity termed the coupling distance \( d_c \) captures the distance-dependent features, with the magnitude of this quantity varying between 2 and 10 Å depending on the residue that is mutated, its location in the protein, and the interactions that it mediates. Mutations can, therefore, subtly influence the folding status and hence dynamics of
residues not just in the immediate vicinity but even as far as 20 Å from the mutation site in a context-dependent manner. The results presented here explain why residues around active sites of proteins (up to 30 Å) exhibit lower evolutionary mutational rates compared to residues farther from the active site (Jack et al., 2016). While mutations can be classified as having an effect on function or not (neutral mutation), the same cannot be said for their effect on modulating populations in the native ensemble that can be highly non-intuitive and challenging to probe from an experimental viewpoint. This observation applies even more so for charged residues on the protein surface that exert a larger effect on surrounding residues. Such modulations aid in the evolution of proteins as the intra-protein interaction network is always perturbed on mutations, in the acquisition and loss of function, and form the basis for epistatic effects.

Limitations of the study

Allostery is a broad field with a rich history, and with multiple experimental and computational treatments contributing to the understanding of this phenomenon (Chen et al., 2022; Grutsch et al., 2016; Lisi and Loria, 2016; Verkhivker et al., 2020; Wodak et al., 2019; Yao and Hamelberg, 2019). In this work, we have highlighted and explained the origins of an oft-observed and an under-appreciated hallmark of mutational effects—the distance-dependent decay of perturbations, particularly those observed in NMR experiments. Exponential-like decays are the zeroth-order expectations from protein structural perturbations, while deviations from this behavior can be attributed to the specific and idiosyncratic nature of protein structures and interaction energies. Specifically, if regions adjacent to the site of mutations are only weakly coupled, the domino-like effect of perturbations could contribute to significant unraveling of the local structural features (for example, melting of helices or enhanced dynamics in loops). In such cases, these residues will not follow the exponential-like pattern observed here—this is particularly expected in case of mutations not just in the immediate vicinity but even as far as 20 Å from the mutation site in a context-dependent manner. The results presented here explain why residues around active sites of proteins (up to 30 Å) exhibit lower evolutionary mutational rates compared to residues farther from the active site (Jack et al., 2016). While mutations can be classified as having an effect on function or not (neutral mutation), the same cannot be said for their effect on modulating populations in the native ensemble that can be highly non-intuitive and challenging to probe from an experimental viewpoint. This observation applies even more so for charged residues on the protein surface that exert a larger effect on surrounding residues. Such modulations aid in the evolution of proteins as the intra-protein interaction network is always perturbed on mutations, in the acquisition and loss of function, and form the basis for epistatic effects.

Limitations of the study

Allostery is a broad field with a rich history, and with multiple experimental and computational treatments contributing to the understanding of this phenomenon (Chen et al., 2022; Grutsch et al., 2016; Lisi and Loria, 2016; Verkhivker et al., 2020; Wodak et al., 2019; Yao and Hamelberg, 2019). In this work, we have highlighted and explained the origins of an oft-observed and an under-appreciated hallmark of mutational effects—the distance-dependent decay of perturbations, particularly those observed in NMR experiments. Exponential-like decays are the zeroth-order expectations from protein structural perturbations, while deviations from this behavior can be attributed to the specific and idiosyncratic nature of protein structures and interaction energies. Specifically, if regions adjacent to the site of mutations are only weakly coupled, the domino-like effect of perturbations could contribute to significant unraveling of the local structural features (for example, melting of helices or enhanced dynamics in loops). In such cases, these residues will not follow the exponential-like pattern observed here—this is particularly expected in case of mutations not just in the immediate vicinity but even as far as 20 Å from the mutation site in a context-dependent manner. The results presented here explain why residues around active sites of proteins (up to 30 Å) exhibit lower evolutionary mutational rates compared to residues farther from the active site (Jack et al., 2016). While mutations can be classified as having an effect on function or not (neutral mutation), the same cannot be said for their effect on modulating populations in the native ensemble that can be highly non-intuitive and challenging to probe from an experimental viewpoint. This observation applies even more so for charged residues on the protein surface that exert a larger effect on surrounding residues. Such modulations aid in the evolution of proteins as the intra-protein interaction network is always perturbed on mutations, in the acquisition and loss of function, and form the basis for epistatic effects.

Limitations of the study

Allostery is a broad field with a rich history, and with multiple experimental and computational treatments contributing to the understanding of this phenomenon (Chen et al., 2022; Grutsch et al., 2016; Lisi and Loria, 2016; Verkhivker et al., 2020; Wodak et al., 2019; Yao and Hamelberg, 2019). In this work, we have highlighted and explained the origins of an oft-observed and an under-appreciated hallmark of mutational effects—the distance-dependent decay of perturbations, particularly those observed in NMR experiments. Exponential-like decays are the zeroth-order expectations from protein structural perturbations, while deviations from this behavior can be attributed to the specific and idiosyncratic nature of protein structures and interaction energies. Specifically, if regions adjacent to the site of mutations are only weakly coupled, the domino-like effect of perturbations could contribute to significant unraveling of the local structural features (for example, melting of helices or enhanced dynamics in loops). In such cases, these residues will not follow the exponential-like pattern observed here—this is particularly expected in case of mutations not just in the immediate vicinity but even as far as 20 Å from the mutation site in a context-dependent manner. The results presented here explain why residues around active sites of proteins (up to 30 Å) exhibit lower evolutionary mutational rates compared to residues farther from the active site (Jack et al., 2016). While mutations can be classified as having an effect on function or not (neutral mutation), the same cannot be said for their effect on modulating populations in the native ensemble that can be highly non-intuitive and challenging to probe from an experimental viewpoint. This observation applies even more so for charged residues on the protein surface that exert a larger effect on surrounding residues. Such modulations aid in the evolution of proteins as the intra-protein interaction network is always perturbed on mutations, in the acquisition and loss of function, and form the basis for epistatic effects.
involving charged residues or perturbations of charged residues located distant from the mutated site. Deviations from exponential-like trends can be studied only on a case-by-case basis and not addressed in this work. Moreover, we have considered only the overall magnitude of the perturbation but not its sign—are the distant residue less coupled (negative $\Delta \Delta G$) or more coupled (positive $\Delta \Delta G$) and why? Answering these questions requires a more detailed analysis where the nature of microstates populated are also studied. Finally, the model considers only native-like energetics and any potential non-native interactions stabilized by mutations are neither predicted and nor accounted for.

**STAR METHODS**

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **METHOD DETAILS**
  - Model description
  - Mutations and DCMs

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.105181.

**ACKNOWLEDGMENTS**

This work was supported by the grant MTR/2019/000392 from Science and Engineering Research Board (SERB, India). A.N.N. acknowledges financial support from the Ministry of Education, New Delhi (Sanction No. 11/9/2019-U.3(A)), and the Centre of Excellence in Biochemical Sensing and Imaging Technologies (CenBioSim), Indian Institute of Technology Madras. A.K. acknowledges funding from the PMRF scheme (Ministry of Education, Government of India).

**AUTHOR CONTRIBUTIONS**

A.N.N. conceptualized the question and design; A.K. performed the calculations; A.N.N. and A.K. analyzed data, prepared figures, and wrote the manuscript.

**DECLARATION OF INTERESTS**

The authors declare no competing interests.

Received: June 30, 2022

Accepted: September 16, 2022

Published: October 21, 2022

**REFERENCES**

Biddle, J.W., Martínez-Corral, R., Wong, F., and Gunawardena, J. (2021). Allosteric conformational ensembles have unlimited capacity for integrating information. Elife 10, e65498. https://doi.org/10.7554/elife.65498.

Bouvignies, G., Vallurupalli, P., Hansen, D.F., Correia, B.E., Lange, O., Bah, A., Vernon, R.M., Dahquist, F.W., Baker, D., and Kay, L.E. (2011). Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. Nature 477, 111–114. https://doi.org/10.1038/nature10349.

Bryngelson, J.D., Onuchic, J.N., Socci, N.D., and Wolynes, P.G. (1999). Funnel pathways, and the energy landscape of protein-folding - a synthesis. Proteins 21, 167–195. https://doi.org/10.1002/prot.340210302.

Buller, A.R., van Roye, P., Cahn, J.K.B., Scheele, R.A., Herger, M., and Arnold, F.H. (2018). Directed evolution mimics allosteric activation by stepwise tuning of the conformational ensemble. J. Am. Chem. Soc. 140, 7256–7266. https://doi.org/10.1021/jacs.8b03490.

Consonni, R., Santomo, L., Fusi, P., Tortora, P., and Zetta, L. (1999). A single-point mutation in the extreme heat- and pressure-resistant sso7d protein from sulfolobus solfataricus leads to a
major rearrangement of the hydrophobic core. Biochemistry 38, 12709–12717. https://doi.org/10.1021/bi9911280.

Cooper, A., and Dryden, D.T. (1984). Allotropy without conformational change. A plausible model. Eur. Biophys. J. 11, 103–109. https://doi.org/10.1007/BF00226625.

Doshi, U., Holliday, M.J., Eisenmesser, E.Z., and Hameleberg, D. (2016). Dynamical network of residue-residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. Proc. Natl. Acad. Sci. USA 113, 4735–4740. https://doi.org/10.1073/pnas.1523573113.

Fersht, A.R., Matouschek, A., and Serrano, L. (1992). The folding of an enzyme 1. Theory of protein engineering analysis of stability and pathway of protein folding. J. Mol. Biol. 224, 771–782. https://doi.org/10.1016/0022-2836(92)90561-W.

Fodor, A.A., and Aldrich, R.W. (2004). On evolutionary conservation of thermodynamic coupling in proteins. J. Biol. Chem. 279, 19046–19050. https://doi.org/10.1074/jbc.M402560200.

Freire, E. (1999). The propagation of binding interactions to remote sites in proteins: analysis of the binding of the monoclonal antibody D1 to lysozyme. Proc. Natl. Acad. Sci. USA 96, 10118–10122. https://doi.org/10.1073/pnas.96.18.10118.

Gerek, Z.N., and Ozkan, S.B. (2011). Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning. PLoS Comput. Biol. 7, e1002154. https://doi.org/10.1371/journal.pcbi.1002154.

Gopi, S., Aranganathan, A., and Naganathan, A.N. (2019). Thermodynamics and folding landscapes of large proteins from a statistical mechanical model. Curr. Res. Struct. Biol. 1, 6–12. https://doi.org/10.1016/j.crstbi.2019.10.002.

Grutsch, S., Brüscheweiler, S., and Tollinger, M. (2016). NMR methods to study dynamic allosterity. PLoS Comput. Biol. 12, e1004620. https://doi.org/10.1371/journal.pcbi.1004620.

Guarrera, E., and Berezovsky, I.N. (2019). On the perturbation nature of allosteric sites: mutations, and signal modulation. Curr. Opin. Struct. Biol. 56, 18–27. https://doi.org/10.1016/j.sbi.2018.10.008.

Harrisonia, A., Verma, R., Purohit, N., Twarog, M.Z., Deshaies, R.J., Bolon, D., and Fushchin, D. (2008). Mutations in the hydrophobic core of ubiquitin differentially affect its recognition by receptor proteins. J. Mol. Biol. 375, 799–996. https://doi.org/10.1016/j.jmb.2007.11.016.

Hecht, M., Bromberg, Y., and Rost, B. (2013). News from the protein mutability landscape. J. Mol. Biol. 425, 3937–3948. https://doi.org/10.1016/j.jmb.2013.07.028.

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. USA 95, 9903–9908. https://doi.org/10.1073/pnas.95.17.9903.

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. https://doi.org/10.1371/journal.pbio.1002452.

Lee, A.L. (2015). Contrasting roles of dynamics in protein allosteroy. NMR and structural studies of CheY and the third PDZ domain from PSD-95. Biophys. Rev. 7, 217–226. https://doi.org/10.1007/s12551-015-0169-3.

Liu, G.P., and Loria, J.P. (2016). Solution NMR spectroscopy for the study of enzyme allostery. Chem. Rev. 116, 6233–6369. https://doi.org/10.1021/acs.chemrev.6b00541.

Naganathan, A.N., and Kannan, A. (2021). A hierarchy of coupling free energies underlie the thermodynamic and functional architecture of protein structures. Curr. Res. Struct. Biol. 3, 257–267. https://doi.org/10.1016/j.crsbi.2021.09.003.

Offenbacher, A.R., Hu, S., Poss, E.M., Carr, C.A.M., Scouras, A.D., Prigogin, D.M., Javaron, A.T., Palla, A., Alber, T., Fraser, J.S., et al. (2017). Hydrogen-deuterium exchange of lipoygenase uncovers a relationship between distal, solvent exposed protein motions and the thermal activation barrier for catalytic proton-coupled electron tunneling. ACS Cent. Sci. 3, 570–579. https://doi.org/10.1021/acscentsi.7b00142.

Pacheco-Garcia, J.L., Anoz-Carbonell, E., Vankova, P., Kannan, A., Palomino-Morales, R., Mesa-Torres, N., Salido, E., Man, P., Medina, M., Naganathan, A.N., et al. (2021). Structural basis of the pleiotropic and specific phenotypic consequences of missense mutations in the multifunctional NAD(P)H:quinone oxidoreductase 1 and their pharmaceutical rescue. Redox Biol. 46, 102112. https://doi.org/10.1016/j.redox.2021.102112.

Petrovic, D., Rasso, V.A., Kamerlin, S.C.L., and Sanchez-Ruiz, J.M. (2018). Conformational dynamics and enzyme evolution. J. R. Soc. Interface 15, 20180330. https://doi.org/10.1098/rsif.2018.0330.

Puri, S., Chen, S.N., Chiu, Y.H., Drackzowski, P., Ko, K.T., Yang, T.J., Wang, Y.S., Uchijima, S., and Hsu, S.T.D. (2022). Impacts of cancer-associated mutations on the structure-activity relationship of BAP1. J. Mol. Biol. 434, 167553. https://doi.org/10.1016/j.jmb.2022.167553.

Rajasekaran, N., and Naganathan, A.N. (2017). A self-consistent structural perturbation approach for determining the magnitude and extent of allosteric coupling in proteins. Biochem. J. 474, 2379–2388. https://doi.org/10.1042/BCJ20170304.

Rajasekaran, N., Sekhar, A., and Naganathan, A.N. (2017a). A universal pattern in the percolation and dissipation of protein structural perturbations. J. Phys. Chem. Lett. 8, 4779–4784. https://doi.org/10.1021/acs.jpclett.7b02021.

Rajasekaran, N., Suresh, S., Gopi, S., Raman, K., and Naganathan, A.N. (2017b). A general mechanism for the propagation of mutational effects in proteins. Biochemistry 56, 294–305. https://doi.org/10.1021/acs.biochem.00079.

Roche, J., Caro, J.A., Dellarole, M., Guca, E., Royer, C.A., Garcia-Moreno, B.E., Garcia, A.E., and Roumestand, C. (2013). Structural, energetic, and dynamic responses of the native state ensemble of staphylococcal nuclease to cavity-creating mutations. Proteins 81, 1069–1080. https://doi.org/10.1002/prot.24231.

Sanchez-Ruiz, J.M., and Makhadzade, G.I. (2001). To charge or not to charge? Trends Biotechnol. 19, 132–135. https://doi.org/10.1016/s0167-7799(00)01548-1.
Tang, Q., and Fenton, A.W. (2017). Whole-protein alanine-scanning mutagenesis of allostery: a large percentage of a protein can contribute to mechanism. Hum. Mutat. 38, 1132–1143. https://doi.org/10.1002/humu.23231.

Tee, W.-V., Guarnera, E., and Berezovsky, I.N. (2020). Disorder driven allosteric control of protein activity. Curr. Res. Struct. Biol. 2, 191–203. https://doi.org/10.1016/j.crstbi.2020.09.001.

The PyMOL Molecular Graphics System, Version 1.2r, Schrödinger, LLC

Tokuriki, N., and Tawfik, D.S. (2009). Stability effects of mutations and protein evolvability. Curr. Opin. Struct. Biol. 19, 596–604. https://doi.org/10.1016/j.sbi.2009.08.003.

Verkhivker, G.M., Agajanian, S., Hu, G., and Tao, P. (2020). Allostery in its many disguises: from theory to applications. Structure 27, 566–578. https://doi.org/10.1016/j.str.2019.01.003.

Xiao, F., Zhou, Z., Song, X., Gan, M., Long, J., Verkhivker, G., and Hu, G. (2022). Dissecting mutational allosteric effects in alkaline phosphatases associated with different Hypophosphatasia phenotypes: an integrative computational investigation. PLoS Comput. Biol. 18, e1010009. https://doi.org/10.1371/journal.pcbi.1010009.

Xu, J., Baase, W.A., Baldwin, E., and Matthews, B.W. (1998). The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. Protein Sci. 7, 158–177. https://doi.org/10.1002/pro.5560070117.

Yang, G., Hong, N., Baier, F., Jackson, C.J., and Tokuriki, N. (2016). Conformational tinkering drives evolution of a promiscuous activity through indirect mutational effects. Biochemistry 55, 4583–4593. https://doi.org/10.1021/acs.biochem.6b00561.

Yao, X.-Q., and Hamelberg, D. (2019). Detecting functional dynamics in proteins with comparative perturbed-ensembles analysis. Acc. Chem. Res. 52, 3455–3464. https://doi.org/10.1021/acs.accounts.9b00485.

Yu, M., Chen, Y., Wang, Z.L., and Liu, Z. (2019). Fluctuation correlations as major determinants of structure- and dynamics-driven allostery effects. Phys. Chem. Chem. Phys. 21, 5200–5214. https://doi.org/10.1039/c8cp07859a.

Zayner, J.P., Antoniou, C., French, A.R., Hause, R.J., Jr., and Sosnick, T.R. (2013). Investigating models of protein function and allostery with a widespread mutational analysis of a light-activated protein. Biophys. J. 105, 1027–1036. https://doi.org/10.1016/j.bpj.2013.07.010.
STAR METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| Deposited data      |        |            |
| Coupling free energies for the three proteins and the alanine mutants | This paper | https://github.com/AthiNaganathan/Mutations-EnsembleEffects |
| Software and algorithms | https://www.mathworks.com/products/matlab.html | R2020a |

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Athi N Naganathan (athi@iitm.ac.in).

Materials availability
The study did not involve experiments.

Data and code availability
- All data are available at the GitHub repository https://github.com/AthiNaganathan/Mutations-EnsembleEffects.
- Any additional information or code required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Model description
We employ the block version of the Wako-Saitó-Muñoz-Eaton (bWSME) model, where a set of contiguous residues are considered as blocks (b), and all the residues in a block can either adopt a folded (represented as 1s) or an unfolded (0s) conformation. (Gopi et al., 2019) A single structure input to the model generated millions of microstates, with each microstate represented by a string of 0s and 1s. For computational efficiency, only microstates that follow the Single Sequence Approximation (SSA, a single island or contiguous stretch of folded blocks), the Double Sequence Approximation (DSA, two islands with intervening unfolded blocks) and DSA with loop (DSA w/L, two islands with interactions considered between them, despite being separated by unfolded blocks) are allowed. The statistical weights of the microstates are estimated considering the contributions from van der Waals interactions with a cut-off distance of 5 Å, electrostatics with no cut-off distance and implicit solvation terms (together included under stabilization free energy) as well as conformational entropy terms for fixing a particular residue in its native state. The rigidity of the proline backbone and the high flexibility of glycine residues are accounted for in the conformational entropy terms. The calculations in this study are performed at 310 K, an effective ionic strength of 0.1 M, and with the pH 7.0 protonation state. The parameters used for the partition function calculation of the three proteins analyzed in this study — PDZ, CheY and CypA— are listed in Table S1.

Mutations and DCMs
Every residue in the wild-type proteins, except alanine, glycine and proline, is mutated to alanine using PyMOL (The PyMOL Molecular Graphics System LLC.), thus resulting in a total of 307 mutations. The structures obtained using PyMOL are fed as an input to create the contact maps for van der Waals interactions and electrostatics of the corresponding mutant. In other words, each mutant has different contact maps, owing to the mutated side chain that alters the interaction network of the protein. These contact maps are used to generate the ‘allowed’ microstates, calculate the total partition function and hence, the probability of every microstate for all the mutants. The microstates are categorized into two
sub-ensembles, and the positive thermodynamic coupling free energies (ΔG+) are calculated (Equation 1) (Naganathan and Kannan, 2021), the details of which are mentioned in the ‘Results and Discussion’ section. Briefly, the summations of probabilities of all the microstates where residue i is folded and those in which i is unfolded, given that residue j is folded, are estimated for the ΔG+ between each pair of residues for all the wild-type and mutant proteins. Note that the ΔG+ matrix obtained is asymmetric as the effect of perturbation of i on j is not the same as j on i due to different environment around the residues. The difference between the pairwise ΔG+ of a mutant (ΔG+,mut) with respect to that of the wild-type (ΔG+,WT) is represented in the form of a ΔΔG+ matrix or a heat map, where the rows and columns indicate the residue number and the color stands for the ΔΔG+ values. This representation is referred to as the Differential Coupling Matrix (DCM). To analyze the distance-dependent effect of mutations, we averaged the absolute values of ΔΔG+ across the rows of the DCMs to obtain the residue-level Differential Coupling Indices (DCIs) in the form of a N×1 vector, N being the number of residues. The distances from the mutated site (Cα-Cα distances) are calculated using the structure of the protein.