An *ab initio* information-theoretic approach to proteins and protein-ligand interactions

Deep Nath,[1] Rajdeep Kaur Grewal,[1] Devrani Mitra,[2,*] and Soumen Roy[1,†]

[1] *Department of Physical Sciences, Bose Institute, Kolkata 700 091, India*

[2] *Department of Life Sciences, Presidency University, Kolkata 700073, India*

Abstract: Differing conformational structure of a protein, associated with two distinct signaling states or between ligand-free and ligand-bound states, leads to differing inter-residue interactions and consequently different biological function. We propose a fresh first-principles information-theoretic approach for studying such proteins and their interactions. A *de novo* measure called protein residue information (PRI), which incorporates details of interactions between all pairs of atoms within and across all residues of the protein, is introduced herein. We formulate a method to calculate the intra-state and inter-state entropy of every residue, needed to determine PRI across any two states of a protein. The intra-state entropy can be determined for every state of a protein possessing one or more states. The inter-state entropy can be calculated pairwise for proteins possessing more than one state. We analyze twenty eight distinct pairs of protein structures from ten different classes. PRI successfully identifies important residues displaying significant conformational changes bearing influence with respect to itself and all other residues. Furthermore, it also successfully identifies important residues displaying rather subtle conformational changes. The identified residues exhibit influential roles in diverse performative features of proteins like stability, allostery, signaling, etc. PRI successfully recovers known experimental results from literature and predicts important roles for many hitherto unstudied residues.

I. INTRODUCTION

Understanding the relation between the structure and function of a protein is a grand aim of researchers in various disciplines. These are known to intimately influence each other. Generally, a protein's structure is rather flexible and undergoes conformational changes leading to diverse functional outcomes. Conformational changes are triggered in the presence of stimuli, ligand binding and other environmental factors. Further, protein folding/misfolding and mutations [1] can also induce conformational fluctuations in a protein. During conformational changes, bonds between the atoms of amino acid sequences remain unchanged and the dihedral angles of the main chain and the orientation of side chains may get altered.

Even seemingly insignificant conformational changes could lead to large or remarkable changes in protein function. Examples range from alteration of catalytic constants of enzymes to negative or positive cooperativity of receptors or even functional alterations in the entire organisms such as swimming patterns of bacteria [2]. Residues can play a crucial role in allostery without undergoing conformational changes [3, 4]. Herein, we formulate an *ab initio* information-theoretic approach to study the consequence of changes in positions of atoms of all residues in a protein. These changes at the atomic-level decide the conformational changes in residues to a varying degree. Such changes, whether small or large, can significantly influence the function of a protein.

Molecular dynamics (MD) simulations provide robust insights into conformational changes of a protein [5, 6]. They deal with the varied dynamical and conformational structures of a protein, to predict associated functional properties. However, they are computationally expensive. It is highly challenging to capture delayed conformational changes on timescales outside the scope of MD simulations [7]. Furthermore, they invoke mathematical approximations to mitigate computational requirements, which in principle, can influence the outcome. Apart from MD, the transitions between various conformational structures could be understood using a protein contact network (PCN) [8, 9]. PCNs help in studying various conformational states of a protein [10]. They find its application in unraveling details of varied phenomena like protein folding, allostery [11], dimer stability [12], residue fluctuations, dynamic structure fluctuations, etc. However, the construction of these network models is non-trivially affected by the choice of the cutoff [8]. Further, these network models often overlook the specific nature of non-covalent interactions (van der Waals interactions or hydrogen bonding) and other complex aspects of the structure of a protein. Further, MD and PCN mostly identify the functional importance of residues, which have undergone significant conformational changes.

The information-theoretic approach proposed herein is plagued neither by demanding computational expense nor by the vagaries of cutoff distances. Moreover, the approach successfully identifies residues that display rather subtle conformational differences between two different states but have significant biological importance. Consider any two given states of the same protein. These states could be experimentally determined structures readily available in the Protein Data Bank (PDB), or obtained through approaches like homology modeling [13] or deep learning, viz. AlphaFold [14]. These structures can represent the protein: (a) in two distinct signaling

* devrani.dbs@presiuniv.ac.in

† soumen@jcbose.ac.in

states, (b) before and after ligand binding, or, (c) in two different conformations under various other stimuli. In response to external stimuli, the position of atoms of all residues in a protein could change leading to a change in position and/or conformation of the residues. However, it must also be remembered that when such changes occur, the surrounding residues often also rearrange themselves to avoid steric clashes so as to make the overall conformation stable. Consequently, the position and/or conformation of every residue with respect to its neighbouring and indeed all other residues also get altered. Both these changes — the deviation of a residue with respect to itself and that of a residue with respect to every other residue — across two structures of a protein, are of similar importance. As elaborated later, the information-theoretic approach presented herein incorporates both these kinds of changes.

Entropy, an outcome of randomness and disorder, imprints its signature across all areas of life. It is extremely useful in understanding many biological phenomena such as protein folding/unfolding, protein-protein interactions [15] and protein-ligand interactions [16]. Various methods have been introduced to calculate the entropy of proteins. Among these, configuration entropy (CE) is used to study interactions and functions of proteins [16, 17]. CE also finds application in MD simulations to study the folding and unfolding of a protein. Additionally, recent developments have enabled the study of conformational changes in protein folding with the help of conformational entropy [18]. There also exist other forms of entropy, which are distance-based. For example, the mathematical expression of entropy in PCN is strongly influenced by the basic criterion of distance-based cutoff during the very construction of PCN. The interaction between residues is effectively a step function and the choice of cutoff dictates the interaction pattern.

We start with experimentally determined macromolecular structures from PDB. Our formulation proposes that the entropy of a protein can be accounted across its states and is independent of any arbitrary cutoff. We can calculate these contributions separately: (a) in any given state of a protein (*"intra-state entropy"*), and (b) across any two given states of a protein (*"inter-state entropy"*). Generally, proteins display conformational change(s) across two given states due to events like signaling or ligand-binding. Due to such changes, the structure of proteins, and the associated entropy will likely differ. Broadly speaking, intra-state entropy can be associated with a particular conformation, while inter-state entropy can be associated with changes in conformation. Intra-state entropy of a residue in a given state is decided by interactions of every atom in that residue with every atom of every other residue. The significance of interactions between residues towards protein-folding and protein stability is well-studied in literature [19, 20]. These interactions are mainly responsible for the conformation of side chains of residues of a protein [20]. On the other hand, inter-state entropy aims to distinguish between the entropy of a residue in a given pair of states. From, these two forms of entropy, we can propose an expression and calculate the *protein residue information* (PRI) of every residue in the protein. By analysing twenty eight distinct pairs of protein structures drawn across ten different classes, we discuss extensively about how each of the residues possessing high PRI essay protein function. Indeed, we provide structure-based mechanistic insights for all residues with higher values of PRI. For well-studied systems, nearly all our findings agree quite well with reported experimental evidence. For the lesser studied systems, we predict hitherto unknown roles of residues, which is amenable to experimental verification.

Our formulation of a distance-dependent form of entropy is inspired by the extremely well-known Lennard-Jones (LJ) potential. LJ potential is very efficient in capturing the nature of interatomic interactions across all sciences. In biology, LJ potential finds its application in soft-docking approach, molecular dynamics simulations and others diverse problems. The use of the LJ potential limits the interaction between any two atoms in an amino acid sequence. This immediately presents the advantage of eliminating arbitrary cut-offs to portray interactions between any two atoms. Unlike MD simulations, herein we deal only with structures of a protein in two different states. Therefore, this technique is also rather computationally inexpensive.

II. MATERIALS AND METHODS

Let $\mathcal{P}$ and $\mathcal{Q}$ denote two *different states* of a protein. $\{\mathcal{R}\}$ and $\{\mathcal{A}\}$ respectively denotes the *set of residues* and the *set of all atoms in all residues*, which are simultaneously present in both $\mathcal{P}$ and $\mathcal{Q}$. Atoms, i and j, in residues, k and l, respectively are separated by the Euclidean distance, d_{ij}^{kl}, where $k, l \in \{\mathcal{R}\}$, and $i, j \in \{\mathcal{A}\}$. The strength of these interatomic interactions between atoms in neighboring residues would expectedly be higher than that between distant residues. Naturally, interatomic interactions would decrease with the increase of d_{ij}^{kl}.
Inspired by well-known Lenard-Jones potential, herein we propose,

$$\omega_{ij}^{kl} = [(\frac{\sigma_i^k}{d_{ij}^{kl}})^6 - (\frac{\sigma_i^k}{d_{ij}^{kl}})^{12}], \tag{1}$$

where, σ_i^k is a constant for a given atom in a residue. ω_{ij}^{kl} is dimensionless. The value of σ_i^k of atom, $i \in \{\mathcal{A}\}$, belonging to residue, $k \in \{\mathcal{R}\}$, is $\sigma_i^k = min(d_{ij}^{kl}) \forall (l, j) \mid l \in \{\mathcal{R}\}, k \neq l; j \in \{\mathcal{A}\}, j \neq i$. $\omega_{ij}^{kl} > 0$ and its value depends only on d_{ij}^{kl}. Between two states, all atoms in all residues might undergo conformational changes. Hence, we consider all possible pairwise combinations of atoms, i.e., both intra-residue and inter-residue, to calculate d_{ij}^{kl}. Details about interactions between atoms within and among residues and the calculation of σ_i^k are depicted in Fig. 1.

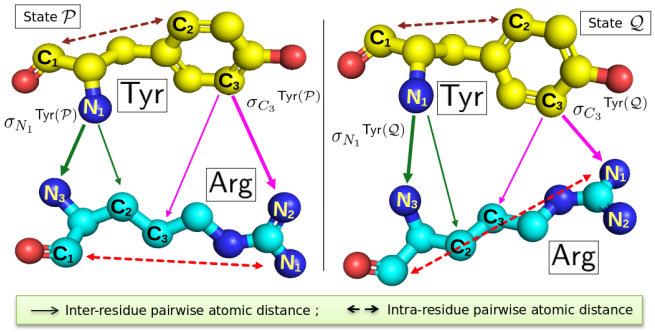

FIG. 1. Illustrative example of two residues of a protein, tyrosine, Tyr, (highlighted in yellow) and arginine, Arg, (highlighted in cyan), undergoing conformational change across the states, $\mathcal{P}$ and $\mathcal{Q}$. Here, we observe that the intra-residue pairwise atomic distances between C_1 and C_2 of Tyr (highlighted in brown dashed line), and, C_1 and N_1 of Arg (highlighted in red dashed line) change between $\mathcal{P}$ and $\mathcal{Q}$. The inter-residue distances of atoms, N_1 and C_3, of Tyr with two different atoms of Arg, depicted in green and magenta respectively, also change across $\mathcal{P}$ and $\mathcal{Q}$. In $\mathcal{P}$, the minimum of all these respective distances of N_1 and C_3 of Tyr with respect to every atom of Arg equals $\sigma_{N_1}{}^{\mathrm{Tyr}(\mathcal{P})}$ and $\sigma_{C_3}{}^{\mathrm{Tyr}(\mathcal{P})}$ respectively. Similarly, $\sigma_{N_1}{}^{\mathrm{Tyr}(\mathcal{Q})}$ and $\sigma_{C_3}{}^{\mathrm{Tyr}(\mathcal{Q})}$ can be determined analogously for $\mathcal{Q}$. $\sigma_{N_1}{}^{\mathrm{Tyr}(\mathcal{P})} \neq \sigma_{N_1}{}^{\mathrm{Tyr}(\mathcal{Q})}$ and $\sigma_{C_3}{}^{\mathrm{Tyr}(\mathcal{P})} \neq \sigma_{C_3}{}^{\mathrm{Tyr}(\mathcal{Q})}$. Generally, $\sigma_i{}^{\mathrm{k}(\mathcal{P})} \neq \sigma_i{}^{\mathrm{k}(\mathcal{Q})}$, for the ith atom of the kth residue of any protein.

Fig. 2 exhibits the conformational change of various residues in two distinct states (holo and apo states) of streptavidin, a biotin-binding protein. As observed in Figs. 2 (a) and 2 (b), the distance between O_1 in Glu51 and N_1 in Arg84 changes from 5.2Å to 5.4Å, implying an increase of only 3.8% across the two states. On the other hand, we observe in Figs. 2 (c) and 2(d), that the distance between O_2 in Glu44 and N_1 in Arg53 across the two states decreases by 61.6% from 7.3Å to 2.8Å. These changes — smaller or larger — in the respective distance between all pairs of atoms across two given states are important to assess the overall conformational changes and its effects, as discussed in Results. Therefore, in our formulation, we consider the most general version — that entropy depends on the distance between all pairs of atoms within and between all residues in a given state and the changes in them across any two states.

A. Entropy

The entropy of a random event, $x_i \in \mathcal{X}$, is related to the amount of information or uncertainty associated with that event. Here $\mathcal{X}$ denotes a complete set of disjoint random events, such that $\bigcup_i x_i = \Omega$, and $x_i \bigcap x_j = \varnothing$. If $p(x_i)$ is the probability of occurrence of event x_i, then the entropy, $\mathcal{S}(\mathcal{X}) = \sum_{x_i \in \mathcal{X}} -p(x_i) \log_2 p(x_i)$. Similarly, if $p(x_i, y_i)$ represents the joint probability of occurrence of two events $x_i \in \mathcal{X}$, and $y_i \in \mathcal{Y}$, then the joint entropy, $\mathcal{J}(\mathcal{X}, \mathcal{Y}) = \sum_{x_i \in \mathcal{X}, y_i \in \mathcal{Y}} -p(x_i, y_i) \log_2 p(x_i, y_i)$. We outline below the procedure for calculation of both intra-state entropy and inter-state entropy.

B. Intra-state entropy

Herein, we evaluate the entropy associated with each individual residue of a protein in a given state. Henceforth, we refer to $\mathcal{S}_k$ as the normalised *"intra-state entropy"* of the kth residue in a given protein. $\mathcal{S}_k^{(\mathcal{P})}$ and $\mathcal{S}_k^{(\mathcal{Q})}$ denote the same for two different states, $\mathcal{P}$ and $\mathcal{Q}$, respectively of the protein. $\mathcal{S}_k^{(\mathcal{P})}$ and $\mathcal{S}_k^{(\mathcal{Q})}$ is decided by the underlying structure of that particular state and can be captured adequately through d_{ij}^{kl} and ω_{ij}^{kl}.

Considering the most general formulation possible, we posit that the probability, p_i^k, of conformational change of the ith atom of the kth residue is influenced in varying measure by all other atoms in every residue excluding the kth residue. Therefore,

$$p_i^k = \frac{\sum_{(j,l)} \omega_{ij}^{kl} (d_{ij}^{kl})^2}{\sum_{i=1}^{N_k} \sum_{(j,l)} \omega_{ij}^{kl} (d_{ij}^{kl})^2}; \quad l \in \{\mathcal{R}\}\backslash\{k\}, j \in \{\mathcal{A}\}\backslash\{i\}. \tag{2}$$

The choice of the functional form, $p^k(\omega_{ij}^{kl}(d_{ij}^{kl})^2)$, is guided by overwhelming experimental evidence and is elaborated in the S.I.

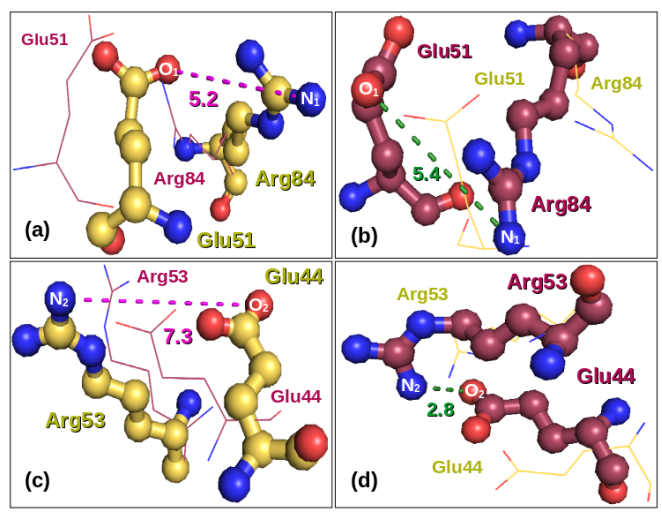

FIG. 2. 1SWE [(a) and (c)] and 1SWA [(b) and (d)] represent the holo (highlighted in golden) and apo (highlighted in maroon) states of streptavidin respectively. The conformation of Glu51 and Arg84 is depicted in (a) and (b), while that of Glu44 and Arg53 is depicted in (c) and (d).The numbers below dotted lines indicate distance in Å.

The corresponding intra-state entropy of the kth residue equals $-\sum_{i=1}^{N_k} p_i^k \log_2(p_i^k)$. Its normalised intra-state entropy therefore is,

$$\mathcal{S}_k = -\sum_{i=1}^{N_k} p_i^k \log_2(p_i^k)/\log_2 N_k, \tag{3}$$

where, N_k, is the total number of atoms present in the kth residue of the protein. A detailed discussion on the physical significance of $\mathcal{S}_k$ with concrete examples can be found in the S.I.

C. Inter-state Entropy

Let $d_{ij}^{kl(\mathcal{P})}$ and $d_{ij}^{kl(\mathcal{Q})}$ denote the pairwise atomic distances in two different states $\mathcal{P}$ and $\mathcal{Q}$ respectively of a protein. The normalised "inter-state entropy", $\mathcal{J}_k^{(\mathcal{P},\mathcal{Q})}$, of the kth residue is calculated *across the states* $\mathcal{P}$ *and* $\mathcal{Q}$, and depends on $\Delta_{ij}^{kl} = d_{ij}^{kl(\mathcal{P})} \sim d_{ij}^{kl(\mathcal{Q})}$ with $k, l \in \{\mathcal{R}\}$ and $i, j \in \{\mathcal{A}\}$.

$\mathcal{J}_k^{(\mathcal{P},\mathcal{Q})}$ is calculated from the structures of two different states of a protein, *and does not arise from direct interactions between atoms of a protein in any given state.* Therefore, physical interactions manifested through ω_{ij}^{kl} and σ_i^k in the case of $\mathcal{S}_k$, are not relevant for the determination of $\mathcal{J}_k^{(\mathcal{P},\mathcal{Q})}$. Similar to the expression of p_i^k used to determine $\mathcal{S}_k$, we posit the probability,

$$q_i^k = \frac{\sum_{(j,l)} (\Delta_{ij}^{kl})^2}{\sum_{i=1}^{N_k} \sum_{(j,l)} (\Delta_{ij}^{kl})^2}, \tag{4}$$

for the determination of inter-state entropy. Thence, we can easily compute the normalised inter-state entropy of the kth residue,

$$\mathcal{J}_k^{(\mathcal{P},\mathcal{Q})} = -\sum_{i=1}^{N_k} q_i^k \log_2(q_i^k)/\log_2 N_k, \tag{5}$$

across states, $\mathcal{P}$ and $\mathcal{Q}$, of the protein. If a given protein has more than two different states, $\mathcal{J}_k^{(\mathcal{P},\mathcal{Q})}$ can be computed *pairwise* $\forall\{\mathcal{P},\mathcal{Q}\}, \mathcal{P} \neq \mathcal{Q}$. Further details can be found in the S.I.

D. Protein Residue Information

Intra-state entropy is associated with the possible interactions experienced by a residue due to other residues. Inter-state entropy incorporates the shift in the position of a residue between two states. To identify functionally important residues, we need to consider the conformation of a residue as well as the interactions it participates in. Therefore, protein residue information (PRI) of residue, k, across two different states, $\mathcal{P}$ and $\mathcal{Q}$, is expressed as,

$$\mathcal{M}_k^{(\mathcal{P},\mathcal{Q})} = \mathcal{S}_K^{(\mathcal{P})} + \mathcal{S}_k^{(\mathcal{Q})} - \mathcal{J}_k^{(\mathcal{P},\mathcal{Q})}, \tag{6}$$

If a given protein has more than two different states, $\mathcal{M}_k^{(\mathcal{P},\mathcal{Q})}$ can be computed *pairwise* $\forall\{\mathcal{P},\mathcal{Q}\}, \mathcal{P} \neq \mathcal{Q}$. As aforementioned, the underlying thought behind PRI is influenced by mutual information (MI). Notionally, MI helps in quantifying the amount of information between two variables, events, or systems. Like MI, $\mathcal{M}_k^{(\mathcal{P},\mathcal{Q})} = \mathcal{M}_k^{(\mathcal{Q},\mathcal{P})}$. However, an interesting difference exists between MI and PRI. While various reports of negative MI in quantum information and entanglement exist [21, 22], it is well-known to be positive in classical physics. $\mathcal{M}_k$ can be negative on occasion, as mentioned in the S.I. This is not unexpected owing to the difference in the physical origin and functional dependence of intra-state and inter-state entropy. To reiterate, $\mathcal{S}_k$ depends on the Euclidean distance between pairs of atoms and is calculated *for any given state*. In contrast, $\mathcal{J}_k$ depends on the *difference* in Euclidean distance between atomic pairs and is calculated *across a pair of states*.

III. RESULTS AND DISCUSSION

We analyze twenty eight distinct pairs of protein structures from ten different classes. Each pair represents two distinct states in PDB, $\mathcal{P}$ and $\mathcal{Q}$. Table I displays all residues possessing $\mathcal{M}_k^{(\mathcal{P},\mathcal{Q})} \geq \langle\mathcal{M}\rangle + (1.5 \times S_{\mathcal{M}})$. The value of $\mathcal{S}_k^{(\mathcal{P})}$, $\mathcal{S}_k^{(\mathcal{Q})}$, $\mathcal{J}_k^{(\mathcal{P},\mathcal{Q})}$, and $\mathcal{M}_k^{(\mathcal{P},\mathcal{Q})}$ of every residue for each of these twenty eight pairs of protein structures is provided in pp. 10-47 of the Supporting Information (S.I.). $\mathcal{M}_k^{(\mathcal{P},\mathcal{Q})}$ denotes the protein residue information of the kth residue. $\langle\mathcal{M}\rangle$ and $S_{\mathcal{M}}$ respectively denote the mean and standard deviation of the distribution of $\mathcal{M}_k^{(\mathcal{P},\mathcal{Q})}$ drawn over all residues of a given protein possessing states $\mathcal{P}$ and $\mathcal{Q}$. We find that not all residues with high PRI values undergo large conformational changes as elaborated in the S.I. This is important because solvent exposed residues with larger side chain could undergo significant conformational changes but bear no structural or functional relevance. On the other hand, some residues often maintain their interactions with the neighbouring residues and are biologically significant. For example, Trp83 in MutY retains its interactions in both states and plays a crucial role in the stability and functionality of MutY, as elaborated later. Our analysis helps in identifying residues that are structurally and functionally important for a protein. Apart from the functional significance of residues with higher values of PRI, we also discuss their conformational changes in great detail in the S.I.

A. CheY

We first investigate CheY, which is a 14KDa response regulator protein well known for chemotaxis and thoroughly studied for its allosteric properties. As plenty of experimental structures are available in PDB, we investigate six different PDB structures in ten different combinations. The structures are: 1JBE (apo CheY in meta-active state – antecedent to fully activated state), 1C4W (D57C mutant), 1F4V (activated, phosphorylated CheY bound to FLiM), 1FQW (activated CheY), 2B1J (unphosphorylated CheY bound to FLiM) and 3CHY

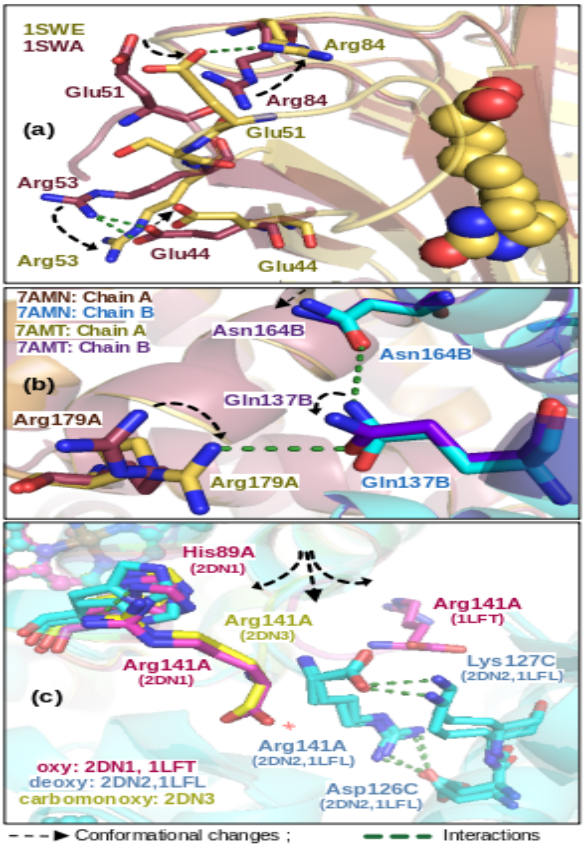

FIG. 3. Formation and destruction of various intra-chain, inter-chain, and inter-subunit interactions in distinct states of: (a) streptavidin, (b) LuxR, and (c) haemoglobin due to conformational changes of residues. (a) Salt-bridge interactions exist between Arg53 and Glu44 in 1SWA. The highlighted contacts break down during the holo-apo (1SWA-1SWE) transition, forming a novel salt-bridge interaction between Glu51 and Arg84 in 1SWE. (b) In 7AMN, there is an intra-chain hydrogen bonding between Asn164B and Gln137B. However, in 7AMT, that interaction is absent and a new inter-chain hydrogen bonding is developed between Gln137B and Arg179A. (c) Inter-subunit hydrogen bonding interactions mediated by Arg141 break down while transitioning from deoxy to ligand-bound states.

(apo-CheY). Three distinct molecular events characterise CheY activation: phosphorylation, conformational changes associated with $\beta4-\alpha4$ loop, and rotational restriction of Tyr106 side chain [23]. While the number and nature of identified amino acids vary in different combinations for obvious reasons, we find an interesting trend. Tyr106 possesses a high value of PRI in all ten combinations of CheY structures as reflected in Table I. Indeed, it is known to play a very crucial role in the activation of CheY [23–25]. It stabilizes the activated configuration of $\beta4-\alpha4$ loop and is present in most of the identified allosteric paths [26]. Experiments designate Thr87-Tyr106 coupling as the main reason behind allosteric effects in CheY [26]. Glu89 appears in all comparisons that contain 2B1J, except one. Indeed in FLiM bound state, Glu89 forms a hydrogen bond with Tyr106, and thus stabilizes the activated state of CheY [23]. Glu89 lies in several predicted allosteric pathways in CheY [25, 26]. Phosphorylation-induced conformational changes of $\beta4-\alpha4$ loop along with four residues Glu89, Trp58, Met85, and Tyr106 results in a buried conformation from solvent-exposed form [27]. Moreover, synchronised movement of both Glu89 and Tyr106 seems extremely relevant to FLiM binding to avoid steric clash with the latter. For the same reason, Lys119 moves farther away to accommodate the C-terminal end of the peptide. Met63, Lys119 and Arg22 appear whenever 1JBE is present in any pair. Met63 appears to change its rotameric state and comes in close contact with the adjacent helix. Close contact of the Met63 side chain with the Leu68 backbone pulls the loop harbouring the phosphorylated Asp57, which eventually allows space to Thr87 and Tyr106 to facilitate FLiM binding. Incidentally, the change in chemical shift for Met63 is observed to be high due to phosphorylation [28]. Though distally located, Arg22 rotates to the opposite direction upon FLiM binding. Change in hydrogen bonding pattern (from β strand-helix contact to intra-helical) is quite distinct in Arg22. This residue is also found to restrict the binding of Mg^{2+} with the structure by influencing the conformation of Asp35 [29]. Further, the synchronised movement of Arg22 and Arg19 (Figure S2(a)) abolishes polar contact between the two in FLiM bound state. Arg19 and Glu34 undergo minor conformational changes between CheY95IV and wild-type structures [30]. With subtle conformational differences, Glu34 seems to stabilise Arg22 in the unliganded state. A combination of experimental results identifies the key role of Ile95 in maintaining the surface complementarity for FLiM binding [30]. In the Ile95Val mutant, besides the Tyr106-Thr87 coupling, Ala88 and Met85 experience significant conformational alterations pertinent to allosteric signaling [27]. $\beta4-\alpha4$ loop (Ala88-Lys91) is known to be responsible for allostery as well [31]. The adjacent Lys92 undergoes minor differences in respect to bond angles when compared between CheY structures solved at 1.1 and 1.7 resolution [23]. However, in the case of CheY-N16-FLiM complex, hydrogen bonding interactions between Lys92 and Asn16-FLiM make the complex stable [25]. Lys109 is an active site residue [24], which behaves as a conformational switch [32]. Phe30 and Phe124 sustain conformational changes, which is triggered by Asp57 due to phosphorylation of CheY [32, 33]. The autophospatase activity of CheY is accelerated manifold in the presence of CheZ. However, Lys26 mutation desists the dephosporylation activity of CheZ [34]. Phospho-CheY (CheY-P) plays a crucial role in chemotaxis, which helps *E. Coli.* to form swarms in agar. A *nonsense mutation* in Lys7 is observed to hinder the swarming ability of *E. Coli* [35]. In apo-state, the side chain of Ser56 is represented as two distinct rotamers [30]. However, it is possible that phosphorylation at adjacent Asp57 perturbs the Ser56 side-chain to move away from Asp57. Glu27 is also observed to have

double conformation in the Ile95Val mutant and its sidechain conformation shows noticeable difference with the wild-type structure [30]. Phe14 located in the vicinity of the phosphorylation site gets buried upon FLiM binding. Four Lysine residues – Lys92, Lys109, Lys119, and Lys126 – are confined to the surface where FLiM, CheZ, and CheA bind [36]. Met129 of *Salmonella enterica* CheY is observed to display high root mean square fluctuation (RMSF) in comparison to *Thermotoga maritima* CheY [37]. Glu93 is located at the solvent-accesible CheA binding surface of CheY [38]. In the activated state, the discrete movement of Glu93 lifts off the steric barrier, thus allowing space for FLiM binding.

B. NolR

NolR is a global transcriptional regulator of genes involved in symbiosis and nodulation, which is necessary for nitrogen fixation. NolR belongs to the winged helix-turn-helix (wHTH) superfamily. Barring a few exceptions, the recognition helix generally binds with the major groove of DNA, while wings comprising of β-turn-β bind with the minor groove. The comparison of 4OMZ-4ON0 enables us to identify critical residues involved in the conformational switching between DNA-bound and unbound forms. Comparative analyses of crystal structures in (un)/liganded forms, thermodynamic analyses, and mutagenesis studies collectively point towards one particular residue, namely Gln56 [39] — as the key player in the NolR mechanism. Conformational switching of this particular residue located at the recognition helix is necessary to accommodate variable target DNA sequences responsible for nodulation and symbiosis. Indeed, our method detects Gln56 as a key residue promoting interactions with nonpalindromic DNA sequences. Incidentally, 80% of the identified residues in both the polypeptide chains of the NolR dimer are located either at the recognition helix or the wing region (Figure S2(b)). Interestingly, our recent findings [40] reveal that 92% of the NolR residues — which exhibit the maximum change in eigenvector centrality — are concentrated in the recognition helix and the wing region. Perhaps this is worth mentioning that the eigenvector centrality measure of protein residue information is effective in identifying key amino acid residues involved in the allosteric signaling process [41]. Gln61 undergoes significant changes in its rotameric form to establish three new contacts with DC18, DT19 of one nucleic acid chain, and DT4 of another chain. Gln61 makes additional contact with conserved Ser57, facilitating bridging interactions between two DNA strands [39]. Similarly, Lys65 too suitably reorients itself to make electrostatic contact with DT17s phosphate group. The guanidium group of Arg67, located at the tip of the recognition helix makes a complete shift to interact with DT8 phosphate. Arg75 and Tyr82 from the wing stabilize the wing structure through hydrogen bonding interactions with the loop residues. Lys12 and Lys40 from different polypeptide chains seem to undergo drastic conformational changes between the two states. They likely have a role in the NolR dimer interface.

C. Streptavidin

Considering its extraordinarily high affinity to biotin (vitamin H), the streptavidin-biotin complex has been hugely used to connect molecules, especially in immunoassays. Herein, we consider the PDB pair 1SWE and 1SWA representing holo-streptavidin and apo-streptavidin respectively to investigate streptavidin-biotin interaction dynamics. In each of the four streptavidin subunits, four Trp residues (at positions 79, 92, 108, and 120) are located at the biotin-binding site. The changes in intrinsic Trp emission fluorescence upon binding with biotin signify positive cooperativity [42]. A comparative study of quenching in Trp emission at 335nm (350nm) regions for hydrophilic (hydrophobic) surroundings identifies the above-mentioned four Trp residues undergoing direct interaction with biotin. Trp92, which is stabilized by stacking interaction from neighbouring Trp75 is one of the critical residues in Trp-tetrad. Replacement of Trp92 by Phe79 in another biotin binding protein Avidin results in altered specificity towards biotin [43]. The higher stability of holo-streptavidin compared to its apo-form has been largely attributed to the movement of Trp120 into the biotin-binding pocket of the neighboring subunit upon binding with biotin. The binding of biotin displaces water molecules from the active site. The eventual burial of biotin takes place through the ordering of the surface loop constituted of residues 45 to 50. Glu51, Ser52 and Arg53 located in subunit C adjacent to the surface loop are instrumental in the apo-holo transition. These residues posess high PRI values. Indeed, a closer examination of this portion further reveals significant structural reorientation in this part of the surface loop (Figure S2(c)). All these residues are completely exposed to the solvent. Glu51C forms a salt bridge interaction with Arg84C in the holo-state, but not in the apo-structure. Conversely Arg53 is salt bridged with Glu44 (preceding the surface loop) in the apostate but not in the holo-state. Upon biotin binding the Arg53-Glu44 salt bridge breaks and instead Arg53 gets completely exposed to the solvent. Selective destruction of one salt-bridge and construction of another salt-bridge at the other end of this three-residue stretch suggests the possible entry of biotin through the Glu51-Arg84 side, as shown in Figure 3(a). In agreement with the experimental data we also find the differential participation of each streptavidin subunit towards biotin binding. To illustrate, Tyr54 makes sidechain-backbone hydrogen bonding/polar interactions with both Glu51-Arg84 in chain C but only with Glu51N in chain A. In chains A and C, the role of Arg84 is found to be more significant for biotin binding. In chains B and D, the adjacent solvent

exposed Tyr83 has enhanced significance. Tyr83 engages in polar interactions with neighboring surface loops, and likely plays a role in biotin entry in the respective subunits. Further, Glu116A (close to Trp120A) interacts with Arg103B (a β barrel residue) via a salt bridge interaction in the holo-form. This particular inter-subunit interaction is not observed between other subunits. Movement of Glu116A towards Arg103B leading to salt bridge formation facilitates biotin binding in subunit A. Despite minor conformational changes, larger changes in the selenoprotein binding to streptavidin is witnessed around Glu116 [44]. Lateral hydrogen bonding interaction between solvent-exposed Tyr22 and Thr20 stabilizes contact regions of higher-order biotinylated streptavidin structures [45]. Val133 also likely has a stabilizing role in the contact regions. Phe130 along with the neighbouring Trp21 from the β barrel provides the necessary hydrophobicity in the biotin-binding pocket.

D. Aureochrome

In the 5A8B-5DKL pair, we investigate the influence of blue light on the conformational dynamics of the signaling aureochrome photoreceptor. Aureochromes [46] are a unique group of blue-light responsive basic leucine zippers (bZIPs), which regulate DNA binding activity upon the reception of light at the light-oxygen-voltage (LOV) domain. Aureo1-LOV from *P. tricornatum* (*Pt*Aureo1 LOV) displays the signature five stranded antiparallel β sheet harboring flavin co-factor, flanked and interspersed with α helices. Packing/unpacking of both the flanking A'_α and J_α helices play a critical role in the conformational reorientation of the dimerization interface upon light induction. The simultaneous unfolding of A'_α, and J_α helices is necessary for light-induced dimerization and transfer of signal to the adjacent bZIP domain for DNA binding activity [47]. The distribution of residues identified from both the polypeptide chains of dimeric *Pt*Aureo1 LOV is shown in Figure S2(d) of SI. Interestingly, none of them is located in the flavin-binding pocket, barring Phe271. Phe271 is proximate to FMN and is fully conserved not just across Aureochromes but throughout the LOV photoreceptor family. Phe46 in *Bacillus sp.* LOV photoreceptor YtvA is homologous to Phe271 in PtAureo1 and undergoes light-induced conformational flip. Mutation of this residue to His significantly accelerates dark recovery kinetics. Conformational flexibility of this residue is known to be important for the dark-light transition [48]. Apart from Phe271, the residues identified herein are dispersed across the LOV structure: (a) featuring mostly on the outer side of β sheet lining dimer interface, (b) on the peptide linkers connecting secondary structure motifs, and (c) on both the terminal helices. Recent studies confirm the role of peptide linkers and C_α helix mediating conformational changes associated with light-induced allosteric signaling [49]. Non-polar interactions between the back of A'_α helix and the β core is crucial to maintain the dark state structure. Following illumination by blue light, the unfolding of A'_α helix exposes the hydrophobic surface of the β core, resulting in reorientation of the dimer interface and thereby affecting signal transduction. Phe241 from A'_α helix, which packs against β strands [50], interacts with Phe332 of the β core in the native dark state. Upon illumination by blue light, the side chain of Phe332 rotates in the opposite direction, breaks its contact with Phe241 and promotes light-induced conformational changes. The phenol ring of Tyr266 also rotates completely in the opposite direction thereby promoting light-induced reorientation of the dimer interface. Again the Tyr277-Tyr320 pair located at the peptide linkers adjacent to C_α and G_β respectively, reorients subtly to make a new polar contact in the light state. This minor shift in distance from 4.4Å to 3.6Å seemingly contributes towards the compaction of the LOV core following the departure of A'_α helix in the light state. A similar effect is noticed in Trp328, which is located between J_α and G_β. Though the polar contact of Trp328 with Asp355 (at J_α) is retained in the light state, the indole ring of Trp328 deviates from planarity and moves farther from J_α. This results in a more compact β core. It is well-known that Asn319 and Asn329 interact directly with the isoalloxazine ring of the flavin chromophore [51] and are conserved components of the robust hydrogen bonding network needed for LOV photochemistry. Remarkably, our studies find the residues adjacent to Asn319 and Asn329, namely Tyr320 and Trp328 as shown in Figure S2(d) of SI. Due to simultaneous unidirectional shifts in the Tyr357-Gln330 pair, hydrogen bonding is retained in between the N-terminal part of J_α and the β core, thereby supporting the partial unfolding of J_α in the light state [51]. Tyr357 further positions itself at the juncture of J_α and A'_α in the light-induced dimer interface due to structural reorientation. Gln366 located at the C-terminal tip of J_α shows a drastic change in conformation between dark-light states. This attests to the unfolding of J_α, which induces light-dependent signaling.

E. LuxR

LuxR is structurally similar to the Tet repressor and belongs to the helix-turn-helix (HTH) group of transcription factors. LuxR is responsible for the activation/repression of hundreds of genes, giving rise to the emergence of quorum sensing in *Vibrio sp.* Therefore, understanding the difference in relative positions of amino acids in LuxR-DNA structures during activation (7AMT) and repression (7AMN), sheds insight into the underlying molecular mechanism leading to such collective behaviorial response. LuxR binds with DNA as a homodimer in both states [52]. A quick PDBePISA [53] analysis interestingly reveals that, while transitioning from activation to repression there is a gross change in protein-protein and protein-DNA contact surface. While the protein-protein interface area expands in activation, the

protein-DNA interface contracts. The opposite happens during repression. Ile199 is one such residue that makes an additional hydrogen bonding interaction with Met196 only in the activation state but not in the repression state. Interchain hydrogen bonding interactions between Asn164A and Tyr193B differ slightly between activation and repression state. Asn164A-Gln176 interaction replaces Asn164A-Tyr193B interaction in the latter state. In the activation state, Asn164 from chain B makes a single hydrogen bonding contact with Tyr193A, whereas Asn164B makes additional interaction with Gln176A besides interacting with Tyr193A in the repression state. Interchain salt bridge interactions involving Arg122 and Glu116 is conserved in both activation and repression state. The chain-wise dynamics of Gln137A (Gln137B), Asn164A (Asn164B) and Arg179B (Arg179A) triad are exciting. From Figure 3(b), we observe that while transitioning from repression to activation state, Gln137A moves subtly to form hydrogen bonding interactions with Asn164A, moving a little farther from Arg179B in the process. However, Gln137B breaks the hydrogen bonding interaction by moving farther from Asn164B, only to create a new hydrogen bonding interaction with Arg179A, as shown in Fig. 3(b). It is indeed interesting to witness antagonistic inter-polypeptide chain interactions as observed in LuxR, which is well-detected by our method. Arg36 and Tyr56 from both A and chain Bs are engaged in hydrogen bonding interactions with both strands of DNA, besides contributing to the protein-protein interaction interface as well. Phe54 demonstrates van der Waals interactions and hydrophobic contacts with T14, A15, and T16 of LuxR-DNA repression complex [52]. During repression, the OG1 atom of Thr59 is at a distance amenable to bonding from the OE1 of Glu61 in contact with the DNA backbone. However, during the activation state this particular Glu61 shifts rather drastically by 4.2Å from the DNA backbone. Further, Thr59 instead of interacting with Glu51, makes polar contact with Asp62, when both of these are at favorable distance from the DNA backbone. Similarly, Arg60 connects with Glu124 in the activation state, whereas during repression, it moves farther away not only from Glu124 but also from the major DNA binding residue Arg36. In fact, DNA binding to LuxR promotes electrostatic interactions in the amino acid tetrad Arg60-Glu124-Arg122-Glu116. While the role of Phe57B has not yet been ascertained precisely, its location at the end of α_3 helix could be important in the context of the neighboring DNA binding residues and the critical dimer interface. The same applies to Phe78B, located distally towards the end of α_4 helix. Phe78 also seems pertinent for maintaining hydrophobicity in the surrounding region. Hydrophobicity in this region has been helpful in designing future quorum sensing inhibitors [54]. Ile93, Glu102, and Gln106 are all located in the α_5 helix of the B subunit. In each case, their side chains move in the opposite direction as part of the overall conformational changes associated with interstate transition. The movement of Ile93B occurs in concert with the movement of Met143B and Phe144B, with both located at the subsequent α_6 helix (Figure S3(a)). Asp65 maintains hydrogen bonding contact with Asn69 in both states. The carbonyl oxygen of Leu87, which is a neighbor to the carbonyl oxygen of Leu85 in the activation model, moves away by $\approx$ 1Å in the repression state. Arg134 moves closer to interact with Ser131 in the repression state but not in the activation state. The most prominent change of Glu152A is the hydrogen bonding interaction between its side chain oxygen with neighboring His88 imidazole nitrogen in the repression state. Interestingly, in the activation state, the His88 side chain rotates by $\approx 90°$, and extends further to lose interaction with Glu152. Remarkably, His88 in chain B remains largely unaffected during the inter-state transition. Asymmetric conformational changes on both the LuxR subunits seem to be the hallmark of its interstate transition. This is further validated by selective rotameric changes in Glu23B, which is not observed in Glu23A. Further, this observation is remarkable in the sense that to switch between activation and repression, even the most distal amino acid residues from the DNA binding region undergo subtle conformational changes and that too asymmetrically.

F. Tbx5

Tbx5 belongs to the T-box transcription factor family and plays a pivotal role in heart and limb development. Several mutations in Tbx5 have been associated with an autosomal disorder called Halt-Oram syndrome. In the 2X6U-2X6V pair [55], we compare the DNA binding domain (DBD, amino acid residue stretch 58-238) of Tbx5 in DNA bound/unbound states and identify twelve residues spanning across the DBD. Among the residues found important in our study, Arg81 and Asn230 interact with DNA. In the DNA bound state (2X6V), the Arg81 side chain extends to establish contact with the phosphate oxygen of DG'18 (Figure S3(b)). In the DNA-bound state, Asn230 flips by 180° not only to place ND2 atom appropriately near the phosphate oxygen of DT'17, but also to place its non-polar backbone at the vicinity of Phe84 from the neighbouring loop. Gln156 and Lys157, both parts of nuclear localization signal (NLS) undergo significant conformational changes during transition to the DNA-bound state. Lys115 exhibits a complete flipping in the DNA-bound state, and belongs to the lid region comprising of β-turn-β. Being located adjacent to DNA binding site, the structural integrity of the lid is extremely important. Glu128 succeeding the lid interacts with Arg113 from the lid in the DNA-free state. However, this polar contact ceases to exist in the DNA-bound state, indicating necessary structural rearrangement in the Tbx5 structure preceding DNA binding. A similar rearrangement is seen for His220. The imidazole chain of His220 flips completely to the opposite direction only to loose contact with another lid-residue Asn119 in the

DNA bound state. Phe57, Tyr100, Arg150 and Gln180 seem to contribute towards structural stability at the region distal from DNA binding site. Another residue with a high PRI value, His170, is crucial for structural stability, as elaborated in the S.I.

G. Hemoglobin

In "polyfunctional" hemoglobin (Hb), we choose the most vital function, namely oxygen transport. The existence of multiple signaling pathways leading to Hb's phenomenal allostery has been confirmed by: (a) multiple ligation intermediates with distinct conformational states and oxygen affinities, and, (b) determination of many structural intermediates through static and dynamic crystallography. In this regard, we compare various PDB pairs - 1LFL vs 1LFT (deoxy vs oxy), 2DN1 vs 2DN2 (oxy vs deoxy), 2DN2 vs 2DN3 (deoxy vs carbomonoxy) and 2DN1 vs 2DN3 (oxy vs carbomonoxy). Comparisons of structure between unliganded tensed state (T) with liganded relaxed state (R) against low and high oxygen affinity respectively provide molecular insights into the cooperative effects in Hb tetramer. Cooperativity among the four subunits of $\alpha_2\beta_2$ tetramer is essential for oxygen uptake and *in vivo* release. Cooperativity is also responsible for shifting the allosteric equilibrium from the T-state to R-state [56]. Structure-function studies supported by computational, thermodynamic, and spectroscopic data confirm the presence of multi-state structures besides the classical T and R end-states, namely $R2$, $RR2$, $R3$ and $RR3$ [56]. As described in the S.I., we identify the residues located across $\alpha\beta$ subunit of Hb. Among these identified residues, Asn102B is crucial as the Asp94A-Asn102B hydrogen bond in the R and $R2$ states is absent in the T-state. His97B undergoes marked differences in its conformation during the $T \rightarrow R$ transition. In fact, the change in interactions mediated by His97B in both T and R states is considered as a major steric barrier to the quaternary structure change. However, the R-$R2$ screw axis passing through the Trp37B-Arg92A intersection induces significant conformational changes in His97B, which is different from the $T \rightarrow R$ switching movement [57]. In the T-state, the C-terminal residue Arg141A interacts electrostatically across α subunits and maintains T-state stability. Figure 3(c) shows that in the R-state, Arg141A flips completely to lose contact with His89A of E-helix. Arg141α, through inter-subunit hydrogen bonding interactions with Asp126α and Lys127α, stabilises the deoxy state and facilitates oxygen release. [56]. Further in the C-terminal region of the T-state, His146B engages in two inter-, and intra- subunit hydrogen bonding interactions. However, given its high solvation and disorderliness, His146 from both the β subunits move drastically to stack against each other (Figure S3(c)) and establish a new salt bridge connection with Lys82B. Lys82B facilitates the $T \rightarrow R$ transition by contributing to Bohr effect via deoxygenation-linked proton binding. Lys82 being a positively charged residue repels other positively charged residues at the central water cavity and increases the free energy of Hb [56]. Lys82 along with His2, Asn139 and His143 are β cleft residues. Interestingly, 2,3 bisphosphoglycerate and its analogues interact with these β cleft residues via water-mediated/direct hydrogen bonding interactions, stabilising T-state structure with low oxygen affinity [56]. In fact in Hb Hinsdale (a variant of human Hb), Asn139 $\rightarrow$ Lys mutation introduces additional positive charge in the central water filled cavity, destabilises T-state structure due to electrostatic repulsion and increases oxygen affinity. In another Hb variant (Hb Doha), Val1 $\rightarrow$ Glu substitution in the β subunit further promotes the T-state structure by preventing the cleavage of the initiator acetylated methionine. This causes structural changes with the formation of intra/inter-subunit contacts especially at the $\beta1\beta2$ interface, resulting in a stable T-state structure [56]. Further, mutations at the $\alpha1\beta1$ or $\alpha2\beta2$ interface can cause significant changes in the oxygen binding properties of Hb. For example, Tyr35 $\rightarrow$ Phe mutation at the β subunit results in increased oxygen affinity and decreased cooperativity in Hb Philly variant. Again, Phe42 deletion causes abrupt shift in allosteric equilibrium, resulting in reduced oxygen affinity and low cooperativity in ligand binding – as found in Hb Bruxelles [58]. Our method detects Thr38 as one of the key residues. Interestingly, Thr38 is located just one helix turn away from Tyr35 and Phe42, and favourably contacts the vinyl group of heme cofactor via its non-polar end of side chain. Further, mutations in the hinge region (e.g. Tyr35 $\rightarrow$ Arg) of $\alpha1\beta2$ perturbs the dimer interface, causing variable affinities towards oxygen [56]. Salt-bridge interactions, especially the ones involving the C-terminal residues of Hb (e.g. Asn139, and Arg104 interaction) play a crucial role in the cooperativity mechanism of HbO_2 [59]. His146 forms two salt bridges in T-state. However, in the R-state, these salt bridges are removed and His146 gets solvated and disordered [57]. In deoxyhemoglobin, His146 gets delocalised and does not participate in salt-bridge formation [59]. His146 also promotes $T \rightarrow R$ transition. Tyr24A and Trp15B were observed to take part in nitration of human hemoglobin by $H_2O_2/nitrite$ [60]. The formation of 4-Hydroxybenzyl adduct at Tyr24A in Hb indicates high solvent accessibility [61]. Arg31 from the B-helix engages in interchain hydrogen bonding interactions with Phe122 backbone and Glu27 side chain respectively from G and B-helices. This facilitates the docking of *Staphylococcus aureus* IsdB onto αHb — showing the former's importance in organic iron acquisition [62]. Lys11A-Ala168B contact is the most conserved interaction at $Isdb^{N1} : Hb$ interface [62]. Lys60A also contributes a hydrogen bond at the $Isdb^{N1} : Hb$ interface [62]. Conserved His72A from the EF corner along with the other conserved histidines (89, 97, 103 and 112) are protonated during the transition from R to T-state [63], with distinct role in the Bohr effect. His143B along with His20A, His112A and His77B

are identified to play a role in countering alkaline Bohr effect. His77B is observed to possess different conformations in different structures [64]. Another conserved histidine involved in Bohr effect (His117B), however gets deprotonated while transitioning from R to T-state [63]. In R-state structure, His117B, and Glu22B engage in weak electrostatic interaction. However, in T-state, they disengage and drift apart by ≈ 8 Å. Carboxymethylation and hydroimidazolone modification are observed on Arg30B in human Hb [65]. Phe71B from the E-helix is located at the frontiers between internal cavities of goblin fold and also known as a mechanically sensitive residue [66]. Research on pyrrole based alkaloids is important not just for their structural and conformational similarities with porphyrin moieties but also for the therapeutic purposes like lowering of blood cholesterol. Docking analysis involving Hb and a pyrrole compound, PyS, reveals the significance of Trp14A in facilitating π-π interaction between Hb and ligand [67]. T-R structure comparison reveals complete flipping of indole side chain of Trp14A. Sodium dithionite is popularly used as deoxygenation agent of Hb and Mb. Crystal structure reveals the interaction of dithionite molecule with Hb via His116B, His117B and Lys16A residues, thereby promoting deoxygenated state [68]. Thr67 is located at the channel connecting heme iron binding site of α subunit of Hb, and the exterior. Substitution of Thr67 with more obstructing but more mobile Lys suggests similar CO exit path in Hb and Mb [69]. Location of Val70A at the channel connecting the distal site and solvent suggests its role in ligand migration [70]. The heme pocket is generally conserved between CO-Hb and CO-Mb. NMR studies reveal conserved hydrogen bonding interaction between the backbone carbonyl of Leu83A and His87 at the proximal site of CO-Hb. This validates an extremely conserved coordination geometry and similar hydrogen bonding patterns between these two proteins [71].

H. Myoglobin

Effective discrimination between oxygen and carbon monoxide by Hb and Mb is a prime factor to aerobic respiration [72]. Therefore, besides Hb we also focus on Mb. We primarily compare five experimental structures of Mb: 1MDN - V68N de-oxy Mb; 1MNO - V68N oxy-Mb; 1M6C - V68N Mb with CO; 1MWD - wild type de-oxy Mb; 1MWC - wild-type Mb with CO (see S.I.). While comparing the aquomet, unliganded, CO and O_2-bound Mb structures, differences in electron density are mostly observed at the ligand binding site as well as the hydration network surrounding the solvent side of distal His (His64). Mutation of His64 to Leu forbids the system from binding with O_2 [73]. In the CO-bound form, His64 swings out of the distal heme pocket while the latter gets occupied by two ordered H_2O molecules. The movement of His64 is in harmony with that of Lys45, which borders the polar heme surface. Lys45 along with His64 and Val68 borders the entry/exit path of diatomic ligand [74]. Further, the mutation of Lys45 by Arg lowers the inner barrier for CO rebinding — as apparent in our findings. After getting dislodged from the CO-binding pocket of Mb, the imidazole side chain of His64 establishes hydrogen bonding/electrostatic contact with heme as well as Asp60. Justifiably, His64 and Asp60 does not feature when wild type Mb with CO bound/unbound forms are compared. Structural superposition of 1MDN, 1M6C and 1MWC in Figure S3(d) shows insignificant movement of His64 in wild type Mb's CO-bound structure. This is contrary to that of the CO-bound form in V68N variant. Again, the side chain of Glu83 flips completely to the opposite direction in CO-bound WT Mb forging new interactions with residues from EF corners. This shift does not take place in the V68N variant, where Glu83 maintains polar contacts with residues from F-helix harbouring proximal Histidine. Upon dissociation of CO, the subtle movement of Lys145 at the H helix by 0.6Å ruptures hydrogen bonding interaction with Thr87 from the F helix. Comparison between unliganded and CO-bound Mb reveals shift of FG corner residue Lys98 towards Lys42 from C helix. Gln8 engages with Lys79 at the EF corner and is found to be important for transiting between ligand bound/unbound conformations. An A helix residue, Lys16 drifts its side chain significantly towards GH corner in O_2 bound state (see S.I.). Glu41 from C, Glu83 from EF, Lys96 and Lys98 from FG and Lys145 from H, which exhibit strong dynamic interaction with the heme group in deoxy-Mb are found to be important for unliganded to liganded state. In fact, Lys145 along with Tyr146 are central mediators of dynamic interactions [75] at the C-terminal segment of the F to H helix. Glu41 from C helix and Lys47 from CD corner, pair up via a salt bridge interaction. However, upon switching to ligand bound states (CO/O_2), the salt bridge ruptures. Notably, in Mb structures, several residues including Gln8, Glu41, Lys47, Glu109, Ser117 and Gln128 have double conformers in one/more state(s). One Glu109 conformer salt bridges with His36, which has greater potential to engage in stacking interactions with Phe106 at the ligand (CO/O_2) bound states. Ser117, at the end of G-helix is completely solvent exposed. Therefore, it would have a role in influencing the hydration network of Mb. Gln128 engages in intense hydrogen bonding interactions with residues from G-helix (Gln116 side chain), GH corner (Phe123 backbone) as well as the H-helix (Ser132 side chain). Together, these constitute the G-GH-H hydrogen bonding network. While several residues are common in information transfer from ligand unbound to bound state, subtle change in single amino acid residue apparently activates additional amino acid centres. This is evident not only from our findings but also from altered CO/O_2 binding affinity between WT and the V68N Mb variant.

TABLE I. Residues possessing a value greater than 1.5 times the standard deviation from the mean, of the distribution of protein residue information for the respective pair of protein structures. Functional associations of each of these are provided in the text. The roles of the vast majority of these residues have already been verified experimentally.

Class of protein structure pair	Pairs of PDB IDs	Identified Residues
CheY (signaling)	1JBE, 1C4W	Arg19A, Arg22A, Ser56A, Met63A, Tyr106A, Lys109A, Lys119A
	1JBE, 1F4V	Arg22A, Phe30A, Met63A, Tyr106A, Lys119A
	1JBE, 1FQW	Arg22A, Met63A, Tyr106A, Lys119A
	1JBE, 2B1J	Lys7A, Arg22A, Lys26A, Glu27A, Met63A, Glu89A, Tyr106A, Lys119A
	1JBE, 3CHY	Lys7A, Arg19A, Arg22A, Ser56A, Met63A, Met85A, Tyr106A, Lys119A
	2B1J, 1C4W	Lys7A, Phe14A, Arg19A, Glu89A, Tyr106A, Lys109A, Lys126A
	2B1J, 1F4V	Phe14A, Arg22A, Phe30A, Trp58A, Glu89A, Tyr106A, Phe124A
	2B1J, 1FQW	Phe14A, Glu34A, Lys92A, Tyr106A, Met129A
	2B1J, 3CHY	Lys26A, Glu89, Ile95A, Tyr106A, Lys126A
	3CHY, 1FQW	Phe14A, Lys92A, Glu93A, Tyr106A
NolR(Transcription)	4OMZ, 4ON0	Lys12A, Gln61A, Lys65A, Arg75A, Tyr82A, Lys12B, Lys40B, Gln56B, Gln61B, Arg67B, Arg75B
Streptavidin (Biotin-binding)	1SWE, 1SWA	Tyr22A, Tyr54A, Trp79A, Arg84A, Trp108A, Glu116A, Trp120A, Trp21B, Tyr22B, Trp75B, Trp79B, Tyr83B, Trp92B, Trp108B, Trp120B, Phe130B, Val133B, Trp21C, Tyr22C, Glu51C, Ser52C, Arg53C, Trp79C,Arg84C, Trp92C, Trp108C, Trp120C, Val133C, Trp21D, Tyr22D, Trp75D, Trp79D, Tyr83D, Trp92D, Trp108D, Trp120D, Phe130D
Aureochrome (LOV)	5A8B, 5DKL	Phe241A, Tyr277A, Tyr320A, Trp328A, Phe332A, Tyr357A, Glu366A, Tyr266B, Phe271B, Tyr277B, Tyr320B, Trp328B, Tyr357B
LuxR (Transcription)	7AMN, 7AMT	Arg36A, Phe54A, Tyr56A, Thr59A, Arg60A, Asp65A, Leu87A, Arg134A, Gln137A, Glu152A, Asn164A, Glu23B, Arg36B, Phe54B, Tyr56B, Phe57B, Phe78B, Ile93B, Glu102B, Gln106B, Arg122B, Met143B, Phe144B, Ile199B
TBX5 (Transcription)	2X6U, 2X6V	Phe57A, Arg81A, Tyr100A, Lys115A, Glu128A, Arg150A, Gln156A, Lys157A, His170A, Arg182A, His220A, Asn230A
Hemoglobin (Oxygen storage / transport)	1LFL, 1LFT	Tyr24A, Arg31A, His72A, Arg92A, Lys139A, Val1B, Glu22B, Arg30B, Tyr35B, Phe71B, His77B, Lys82B, Asn102B, Arg104B, His117B, His146B
	2DN1, 2DN2	Trp14A, Lys16A, Tyr24A, Arg31A, Lys127A, Arg141A, Trp15B, Glu22B, Tyr35B, Trp37B, Phe71B, His97B, Arg104B, His146B
	2DN1, 2DN3	Trp14A, Tyr24A, Arg30A, Arg31A, Lys60A, Thr67A, Val70A, His72A, Leu83A, Arg141A, Glu22B, His117B, His143B, His146B
	2DN2, 2DN3	Lys16A, Tyr24A, Arg31A, Lys127A, Arg141A, His2B, Trp15B, Glu22B, Tyr35B, Trp37B, Lys82B, His97B, Arg104B, His146B
Myoglobin (Oxygen storage / transport)	1MDN, 1M6C	Gln8A, Glu41A, Lys47A, His64A, Lys79A, Lys98A, Glu109A, Ser117A, Gln128A, Lys145A
	1MDN, 1MWC	Gln8A, Lys16A, Glu41A, Lys47A, Glu83A, Lys96A, Lus98A, Glu109A, Ser117A, Gln128A, Lys145A
	1MWD, 1M6C	Gln8A, Lys47A, Asp60A, His64A, Lys79A, Glu83A, Lys96A, Lys98A, Glu109A, Ser117A, Gln128A, Lys145A
	1MWD, 1MDN	Gln8A, Lys16A, Glu41A, Lys45A, Lys47A, Lys79A, Glu83A, Lys96A, Lys98A, Glu109A, Ser117A, Gln128A, Lys145A
	1MWD, 1MWC	Gln8A, Glu41A, Lys47A, Lys79A, Glu83A, Lys96A, Lys98A, Glu109A, Ser117A, Gln128A, Ser132A, Lys145A
MutY (Hydrolase)	1RRS, 1RRQ	Trp20A, Trp30A, Gln48A, Glu59A, Leu71A, Glu76A, Trp83A, Leu86A, Glu101A, Asn146A, Gln211A, Glu223A, Leu244A, Trp266A, Leu282A, Trp312A, Arg332A, Trp352A, Arg353A
KcsA (Membrane)	6W0D, 6W0F	Arg64C, Trp68C, Tyr78C, Trp87C, Arg89C, Trp113C, Phe114C, Glu120C
	7MHR, 7MK6	Ala28C, Tyr45C, Trp67C, Trp68C, Tyr78C, Trp87C, Arg89C, Trp113C, Phe114C
	1K4C, 1K4D	Trp26C, Leu35C, Trp68C, Ser69C, Arg89C, Glu118C, Gln119C, Arg121C, His124C

I. MutY

Our next target is the adenine DNA glycosylase, MutY. It functions in the base excision repair (BER) pathway and protects our DNA from potentially genotoxic 8-oxoguanine (8-OG). MutY initiates DNA repair by removing adenine base opposite 8-OG and generates an abasic site (see S.I.). Herein, we compare two structures from PDB: 1RRQ (with damaged DNA bearing 8-OG) and 1RRS (with DNA containing an abasic site). This means that a comparison between these two structures should shed insight into the precise enzymatic action of MutY in the first step of BER. As detailed in the S.I., the amide bearing side chain of Gln48, a residue with high PRI value, interacts with the neighbouring $3'$ base as well as hydrogen bonding interactions with the phosphate backbone [76]. The Watson-Crick face of 8-OG is further stabilised by the backbone amide carbonyl of Gln48 as well as the hydroxyl side chain of Thr49, both interconnected via hydrogen bonding interactions. Complete flipping of the Glu223 side chain from the exposed to the buried state happens simultaneously with a major shift in the DNA strand at DC'16, DA'17, adjacent to the abasic site (DA'18). Glu59 side chain rotates by $\approx 90^\circ$ to be in close contact with *Ile55/O*. The backbone carbonyl of Leu86 further stabilises OG at the minor groove via hydrogen bonding interactions. While the indole nitrogen of Trp30 is close to the amine group of flipped adenine base, the bulky hydrophobic side chain of Trp30 is part of several other hydrophobic amino acids lying in close contact with the adenine base. Trp20, located at the iron-sulfur cluster containing N-terminal

catalytic domain seems to be important for structural integrity of the cluster. Through hydrophobic interactions, Trp20 stabilises Val197, located next to Fe-S coordinating Cys198. Maintaining the functional integrity of FeS cluster is extremely important because change in oxidation state between $[4Fe-4S]^{2+}$ and $[4Fe-4S]^{3+}$ results in the donation of an electron to the damaged site. In absence of FeS cluster, MutY is unable to bind to the DNA, and thus unable to repair OG lesion. $[4Fe-4S]$ cluster is connected to the polypeptide chain via four cysteine residues - 198, 205, 208 and 214. Inversion of $CONH_2$ side chain of Gln211 is also important to maintain contact with Cys205 in the repaired state. The backbone carbonyl of Gln211 interacts with the backbone amide of Cys214 via hydrogen bonding in both states, thereby signifying its role in the FeS cluster domain. This is in agreement with our formulation that even the residues with subtle to almost no conformational changes during transition plays important role in protein function. Other such residues are Trp83 and Trp266. With no visible conformational changes, Trp83 remains connected to the backbone carbonyl of Gly87 in both states (Figure S4(a)). Maintaining proper conformation in this region is absolutely critical not only to stabilise the Hoogsteen face of OG through Tyr88-OH located next to Gly87 but also to facilitate catalytic process mediated by Tyr88-water-Gln48. While the negatively charged phosphate of OG is stabilised by the backbone amide of Leu262, the hydrophobic side chain of the latter is further stabilised by the non-polar part of the indole of Trp312. Trp266, with its bulky non-polar side chain, pushes Arg332 to move closer to static Tyr342-OH to make additional hydrogen bonding interactions in the repaired state. Loss of contact with the backbone of neighbouring Tyr331 induces the Val327 side chain to rotate in the direction opposite to the repaired state. This results in a subtle widening of the loop. Appearance of Leu282 could be an artefact given the loss of electron density connecting Gly281 and Leu282 in both states. The side chain of Leu71 rotates only to make a new non-polar contact with the carbon atom of imizadole ring in His96, located at the neighbouring helix at the repaired state. Further, the imidazole nitrogen of His96 gets connected to Glu76 in the same state, located in the adjacent helix of Leu71. Non-covalent interactions among Leu71-His96-Glu76 triad that form only in the repaired state, bring neighbouring helices close to each other, resulting in gross conformational changes. Similarly, two beta strands move closer to each other through newly developed interactions with Leu244 in the repaired state. The importance of Glu101 is not readily understandable, especially given the partially unresolved neighbouring Arg105 in the damaged state. In the repaired state, Glu101 makes a salt bridge interaction with Arg105. A shift of the Asn146 side chain in harmony with the formation of an abasic site is quite remarkable. As soon as the abasic site is generated, the amide side chain in Asn146 rotates to stabilise dislocated sugar-phosphate backbone of DA'18. Trp352 seems to maintain local hydrophobicity opposite to the DNA binding face — especially by non-polar interaction with Leu269 and π-π interaction with neighbouring Phe268. In contrast, on the DNA facing side, the exposed polar side chain of Arg353 along with Arg350, Glu357 etc. creates necessary polarity for solvent exposure and DNA binding.

J. KcsA

Potassium ion (K^+) channels facilitate the selective transport of K^+ ions from an intracellular to extracellular environment at an extraordinarily high speed. This is essential for maintaining K^+/Na^+ balance in living systems, thus sustaining homeostasis. Herein, we investigate the bacterial K^+ channel KcsA in two different states: (a) conductive or open state, and, (b) relaxed or closed state. For a better understanding, we compare three PDB pairs, albeit in a single polypeptide format (i.e. chain C vs chain C in PDBs of the two states, due to the lack of tetramer data). The pairs of PDBs considered here are: 6W0F-6W0D, 7MHR-7MK6 and 1K4C-1K4D. They all represent S5-S6 segment of KcsA K^+ channel from *Streptomyces lividans*, solved in open and closed state conformations, crystallized under different salt conditions. Incidentally, all of these structure-pairs are solved in an identical space group, thus suitable for structure comparisons. From Table I, we observe that Trp68, Arg89 and Phe114 are a constant presence in all cases, while Tyr78, Trp87, Trp113 are common in the first two pairs. All the residues Arg64, Trp67, Trp68, Ser69 – flank the conserved filter region which is responsible for regulating selective yet high-speed movement of K^+ ions across the membrane (see S.I.). Though not in WT, Trp67 in KcsA E71V mutant structure adopts a different rotamer conformation – this justifies the detection of Trp67 only in 7MHR-7MK6 structure-pair (Figure S4(b)). A 180° rotation of the indole moiety of Trp67 in 7MHR-7MK6 abolishes hydrogen bonding interaction with the neighboring Asp80 upon gate opening. Trp87 is located at the N-terminal end of S6 (see S.I.) and shows significant movement in its side chain upon the opening or closing of the KcsA channel. A comparison of structure across the open and closed state KcsA reveals translation and counter-clockwise rotation of the inner helices (S5, and S6) with respect to the centre of symmetry of the channel. While Trp26 and Ala28 are present at the N-terminal end of S5, Leu35 showing signs of conformational mobility is located around the middle of S5. Tyr45 towards the C-terminal end of S5 shows a nearly perpendicular shift in its phenolate ring – which seems to have a role towards interaction between the pore helix and the adjacent voltage sensor related to channel opening. Fully conserved Tyr78 is a part of the signature sequence of the selectivity filter. The backbone carbonyl oxygen of Tyr78 participates in K^+ coordination, whereas the protein core-facing side chain provides the necessary geometric constraints to maintain the geometry of K^+ binding

sites. While Trp87 faces the inner side of the protein core, the polar side chain of Arg89 faces the intrasubunit space at the extracellular side. Located in the vicinity of the selectivity filter, Arg89 is connected with the tetramer stability of KcsA [77]. Quite interestingly, the conserved residue Gly99 associated with opening and closing of the gate is the neighbouring residue located at the adjacent inner helix, S6. Considering the direct dependence of $K+$ ion conduction and bending of S6 at Gly99 leading to opening of the gate, contribution from the neigbouring residues appear significant. Trp113, Phe114, Glu118, Gln119, Glu120, Arg121 and His124 – are all present in the C-terminal half of the S6 helix (see S.I.). While the non-polar side chain confers structural stability due to its facing the hydrophobic side, the polar side chains provide a more favourable environment for polar K^+ ions and water at the intracellular side. Existence of polar amino acid residues at the end of S6 justifies easy solvent access even in the middle of the transmembrane portion of the channel. A significantly long aqueous passageway followed by a short selectivity filter is the hallmark of KcsA K^+ channel [78]. This explains the very high conductance rate of K^+ ions through KcsA channels.

In summary, we identify and discuss the functional significance of residues with higher values of PRI. Not infrequently, we find that the residues neighbouring an individual residue — rather than the residue in question — rearrange themselves to assist a specific biological activity. We also predict the likely functional significance of other residues identified by our approach, which are amenable to experimental validation.

IV. CONCLUSION

Substantial correlations exist between the structure and function of proteins. The position and conformation of amino acid residues may be altered due to diverse factors. Conformational changes between two states of a protein arise from alterations in position and/or conformation of every residue with: (A) respect to itself, (B) its neighbouring residues, and, (C) indeed all other residues. Common visualisation software enable us to visualise structural alignment of the types A and B. However, all these conformational changes (types A, B, and C) are important as they lead to the construction and destruction of various non-covalent interactions, which influence the overall functioning of a protein. Herein, we introduce a fresh *ab initio* information-theoretic approach to study proteins and protein-ligand interactions, which accommodates all the aforementioned types of structural changes (A, B, and C).

We formulate a method to calculate "inter-state" entropy, $\mathcal{S}_k$, and "intra-state" entropy, $\mathcal{J}_k$. $\mathcal{S}_k$ depends on the Euclidean distance between pairs of atoms and is calculated *for any given state*. $\mathcal{J}_k$ depends on the *difference* in Euclidean distance between pairs of atoms and is calculated *across a pair of states*. We also propose a *de novo* measure called protein residue information (PRI), using $\mathcal{S}_k$ and $\mathcal{J}_k$. The utility of this approach is readily visible as we employ PRI to analyze twenty eight distinct pairs of protein structures drawn from ten different classes. PRI is largely successful in pinpointing residues of structural and functional significance. Remarkably, it successfully identifies important residues, irrespective of the magnitude of conformational changes undergone by them. Besides revalidating functionally important residues cited in the literature, we furnish mechanistic and functional insights for the rest of the residues from the $3D$ structures. The identified residues are found to be functionally important in allostery, structural stability and protein-ligand interactions. MD simulations and network-based techniques are powerful and established techniques. The present approach — even with any attendant handicap — fares better in terms of lower computation time and circumvention of cutoff distances when compared to MD simulations and network-based techniques respectively. Therefore, this formulation can serve as a fresh alternative theoretical approach to detect the contribution and importance of amino-acid residues in protein function.

[1] E. H. Kellogg, A. Leaver-Fay, and D. Baker, Role of conformational sampling in computing mutation-induced changes in protein structure and stability, Proteins: Structure, Function, and Bioinformatics **79**, 830 (2011).

[2] D. E. Koshland, Conformational changes: how small is big enough?, Nature medicine **4**, 1112 (1998).

[3] A. Cooper and D. Dryden, Allostery without conformational change: a plausible model, European Biophysics Journal **11**, 103 (1984).

[4] R. Nussinov and C.-J. Tsai, Allostery without a conformational change? Revisiting the paradigm, Current opinion in structural biology **30**, 17 (2015).

[5] M. Karplus and G. A. Petsko, Molecular dynamics simulations in biology, Nature **347**, 631 (1990).

[6] J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw, Long-timescale molecular dynamics simulations of protein structure and function, Current opinion in structural biology **19**, 120 (2009).

[7] T. J. Lane, D. Shukla, K. A. Beauchamp, and V. S. Pande, To milliseconds and beyond: challenges in the simulation of protein folding, Current opinion in structural biology **23**, 58 (2013).

[8] K. R. Grewal and S. Roy, Modeling proteins as residue interaction networks, Protein & Peptide Letters **22**, 923 (2015).

[9] L. Di Paola, M. De Ruvo, P. Paci, D. Santoni, and A. Giuliani, Protein contact networks: An emerging paradigm in chemistry, Chemical Reviews **113**, 1598 (2013).

[10] R. Kaur Grewal, D. Mitra, and S. Roy, Mapping networks of light-dark transition in LOV photoreceptors, Bioinfor-

matics **31**, 3608 (2015).

[11] L. D. Paola and A. Giuliani, Protein contact network topology: a natural language for allostery, Current Opinion in Structural Biology **31**, 43 (2015).

[12] M. Khamaru, D. Nath, D. Mitra, and S. Roy, Assessing combinatorial diversity of aureochrome bZIPs through genome-wide screening, Cells Tissues Organs (2022).

[13] M. A. Martí-Renom, A. C. Stuart, A. Fiser, R. Sánchez, F. Melo, and A. Šali, Comparative protein structure modeling of genes and genomes, Annual Review of Biophysics **29**, 291 (2000).

[14] H. K. Wayment-Steele, A. Ojoawo, R. Otten, J. M. Apitz, W. Pitsawong, M. Hömberger, S. Ovchinnikov, L. Colwell, and D. Kern, Predicting multiple conformations via sequence clustering and alphafold2, Nature **625**, 832 (2024).

[15] G. P. Brady and K. A. Sharp, Entropy in protein folding and in protein-protein interactions, Current Opinion in Structural Biology **7**, 215 (1997).

[16] C. A. Chang, W. Chen, and M. K. Gilson, Ligand configurational entropy and protein binding, Proceedings of the National Academy of Sciences **104**, 1534 (2007).

[17] M. Karplus, T. Ichiye, and B. Pettitt, Configurational entropy of native proteins, Biophysical journal **52**, 1083 (1987).

[18] R. L. Baldwin and G. D. Rose, Molten globules, entropy-driven conformational change and protein folding, Current Opinion in Structural Biology **23**, 4 (2013).

[19] M. Gromiha and S. Selvaraj, Importance of long-range interactions in protein folding, Biophysical Chemistry **77**, 49 (1999).

[20] G. Faure, A. Bornot, and A. G. de Brevern, Protein contacts, inter-residue interactions and side-chain modelling, Biochimie **90**, 626 (2008), recent advances in complete genome analysis.

[21] M. Horodecki, J. Oppenheim, and A. Winter, Partial quantum information, Nature **436**, 673 (2005).

[22] B. Bertini, K. Klobas, and T.-C. Lu, Entanglement Negativity and Mutual Information after a Quantum Quench: Exact Link from Space-Time Duality, Phys. Rev. Lett. **129**, 140503 (2022).

[23] M. Simonovic and K. Volz, A distinct meta-active conformation in the 1.1-A resolution structure of wild-type apochey, Journal of Biological Chemistry **276**, 28637 (2001).

[24] J. M. Mottonen, D. J. Jacobs, and D. R. Livesay, Allosteric response is both conserved and variable across three chey orthologs, Biophysical Journal **99**, 2245 (2010).

[25] S.-Y. Lee, H. S. Cho, J. G. Pelton, D. Yan, R. K. Henderson, D. S. King, L.-s. Huang, S. Kustu, E. A. Berry, and D. E. Wemmer, Crystal structure of an activated response regulator bound to its target, Nature Structural & Molecular Biology **8**, 52 (2001).

[26] J. Wang, A. Jain, L. R. McDonald, C. Gambogi, A. L. Lee, and N. V. Dokholyan, Mapping allosteric communications within individual proteins, Nature Communications **11**, 1 (2020).

[27] S. J. McDonald, P. Middleton, T. Dowswell, and P. S. Morris, Effect of timing of umbilical cord clamping of term infants on maternal and neonatal outcomes, Evidence-Based Child Health: A Cochrane Review Journal **9**, 303 (2014).

[28] D. F. Lowry, A. F. Roth, P. B. Rupert, F. W. Dahlquist, F. J. Moy, P. J. Domaille, and P. Matsumura, Signal transduction in chemotaxis. A propagating conformation change upon phosphorylation of CheY, Journal of Biological Chemistry **269**, 26358 (1994).

[29] P. Cronet, L. Bellsolell, C. Sander, M. Coll, and L. Serrano, Investigating the structural determinants of the p21-like triphosphate and mg2+ binding site, Journal of Molecular Biology **249**, 654 (1995).

[30] M. Schuster, R. Zhao, R. B. Bourret, and E. J. Collins, Correlated switch binding and signaling in bacterial chemotaxis, Journal of Biological Chemistry **275**, 19752 (2000).

[31] L. Ma and Q. Cui, Activation mechanism of a signaling protein at atomic resolution from advanced computations, Journal of the American Chemical Society **129**, 10261 (2007).

[32] R. B. Bourret, S. K. Drake, S. A. Chervitz, M. I. Simon, and J. J. Falke, Activation of the phosphosignaling protein CheY II. Analysis of activated mutants by 19F NMR and protein engineering, Journal of Biological Chemistry **268**, 13089 (1993).

[33] S. K. Drake, R. B. Bourret, L. A. Luck, M. I. Simon, and J. J. Falke, Activation of the phosphosignaling protein CheY I. Analysis of the phosphorylated conformation by 19F NMR and protein engineering, Journal of Biological Chemistry **268**, 13081 (1993).

[34] M. G. Sanna, R. V. Swanson, R. B. Bourret, and M. I. Simon, Mutations in the chemotactic response regulator, CheY, that confer resistance to the phosphatase activity of CheZ, Molecular Microbiology **15**, 1069 (1995).

[35] K. C. Boesch, R. E. Silversmith, and R. B. Bourret, Isolation and characterization of nonchemotactic CheZ mutants of Escherichia coli, Journal of Bacteriology **182**, 3544 (2000).

[36] J. Yan, R. Barak, O. Liarzi, A. Shainskaya, and M. Eisenbach, In vivo acetylation of CheY, a response regulator in chemotaxis of Escherichia coli, Journal of Molecular Biology **376**, 1260 (2008).

[37] M. Paul, M. Hazra, A. Barman, and S. Hazra, Comparative molecular dynamics simulation studies for determining factors contributing to the thermostability of chemotaxis protein CheY, Journal of Biomolecular Structure and Dynamics **32**, 928 (2014).

[38] D. Shukla and P. Matsumura, Mutations leading to altered CheA binding cluster on a face of CheY, Journal of Biological Chemistry **270**, 24414 (1995).

[39] S. G. Lee, H. B. Krishnan, and J. M. Jez, Structural basis for regulation of rhizobial nodulation and symbiosis gene expression by the regulatory protein nolr, Proceedings of the National Academy of Sciences **111**, 6509 (2014).

[40] A. Boral and D. Mitra, Heterogeneity in winged helix-turn-helix and substrate DNA interactions: Insights from theory and experiments, Journal of Cellular Biochemistry **124**, 337 (2023).

[41] C. F. A. Negre, U. N. Morzan, H. P. Hendrickson, R. Pal, G. P. Lisi, J. P. Loria, I. Rivalta, J. Ho, and V. S. Batista, Eigenvector centrality for characterization of protein allosteric pathways, Proceedings of the National Academy of Sciences **115**, E12201 (2018).

[42] M. J. Waner, J. M. Hiznay, A. T. Mustovich, W. Patton, C. Ponyik, and D. P. Mascotti, Streptavidin cooperative allosterism upon binding biotin observed by differential changes in intrinsic fluorescence, Biochemistry and Biophysics Reports **17**, 127 (2019).

[43] O. Livnah, E. A. Bayer, M. Wilchek, and J. L. Sussman, Three-dimensional structures of avidin and the avidin-biotin complex., Proceedings of the National Academy of Sciences **90**, 5076 (1993).

[44] E. Ayan, B. Yuksel, E. Destan, F. B. Ertem, G. Yildirim, M. Eren, O. M. Yefanov, A. Barty, A. Tolstikova, G. K. Ketawala, *et al.*, Cooperative allostery and structural dynamics of streptavidin at cryogenic-and ambient-temperature, Communications biology **5**, 1 (2022).

[45] O. Azzaroni, M. Mir, and W. Knoll, Supramolecular architectures of streptavidin on biotinylated self-assembled monolayers. Tracking biomolecular reorganization after bioconjugation, The Journal of Physical Chemistry B **111**, 13499 (2007).

[46] A. Deb, R. K. Grewal, S. Roy, and D. Mitra, Residue interaction dynamics in Vaucheria aureochrome1 light-oxygen-voltage: Bridging theory and experiments, Proteins: Structure, Function, and Bioinformatics **88**, 1660 (2020).

[47] E. Herman and T. Kottke, Allosterically regulated unfolding of the $a'\alpha$ helix exposes the dimerization site of the blue-light-sensing aureochrome-lov domain, Biochemistry **54**, 1484 (2015).

[48] A. Möglich and K. Moffat, Engineered photoreceptors as novel optogenetic tools, Photochemical & photobiological sciences **9**, 1286 (2010).

[49] H. Tian, F. Trozzi, B. D. Zoltowski, and P. Tao, Deciphering the allosteric process of the Phaeodactylum tricornutum Aureochrome 1a LOV domain, The Journal of Physical Chemistry B **124**, 8960 (2020).

[50] A. Banerjee, E. Herman, T. Kottke, and L.-O. Essen, Structure of a native-like aureochrome 1a lov domain dimer from phaeodactylum tricornutum, Structure **24**, 171 (2016).

[51] U. Heintz and I. Schlichting, Blue light-induced LOV domain dimerization enhances the affinity of Aureochrome 1a for its target DNA sequence, Elife **5**, e11860 (2016).

[52] J. Zhang, B. Liu, D. Gu, Y. Hao, M. Chen, Y. Ma, X. Zhou, D. Reverter, Y. Zhang, and Q. Wang, Binding site profiles and n-terminal minor groove interactions of the master quorum-sensing regulator luxr enable flexible control of gene activation and repression, Nucleic Acids Research **49**, 3274 (2021).

[53] E. Krissinel and K. Henrick, Inference of Macromolecular Assemblies from Crystalline State, Journal of Molecular Biology **372**, 774 (2007).

[54] B. S. Kim, S. Y. Jang, Y.-j. Bang, J. Hwang, Y. Koo, K. K. Jang, D. Lim, M. H. Kim, and S. H. Choi, Qstatin, a selective inhibitor of quorum sensing in Vibrio species, MBio **9**, e02262 (2018).

[55] C. U. Stirnimann, D. Ptchelkine, C. Grimm, and C. W. Müller, Structural basis of TBX5-DNA recognition: The T-Box domain its DNA-bound and unbound form, Journal of Molecular Biology **400**, 71 (2010).

[56] M. H. Ahmed, M. S. Ghatge, and M. K. Safo, Hemoglobin: structure, function and allostery, Vertebrate and invertebrate respiratory proteins, lipoproteins and other body fluid proteins **94**, 345 (2020).

[57] M. M. Silva, P. H. Rogers, and A. Arnone, A third quaternary structure of human hemoglobin a at 1.7 A resolution., Journal of Biological Chemistry **267**, 17248 (1992).

[58] N. Griffon, C. Badens, D. Lena-Russo, J. Kister, J. Bardakdjian, H. Wajcman, M. C. Marden, and C. Poyart, Hb Bruxelles, deletion of pheβ42, shows a low oxygen affinity and low cooperativity of ligand binding, Journal of Biological Chemistry **271**, 25916 (1996).

[59] B. Shaanan, Structure of human oxyhaemoglobin at 2.1A resolution, Journal of Molecular Biology **171**, 31 (1983).

[60] H.-J. C. Chen, C.-M. Chang, W.-P. Lin, D.-L. Cheng, and M.-I. Leong, H_2O_2/Nitrite-induced post-translational modifications of human hemoglobin determined by mass spectrometry: Redox regulation of tyrosine nitration and 3-nitrotyrosine reduction by antioxidants, ChemBioChem **9**, 312 (2008).

[61] A. T. Rajczewski, L. Ndreu, S. S. Pujari, T. J. Griffin, M. Å. Törnqvist, I. Karlsson, and N. Y. Tretyakova, Novel 4-hydroxybenzyl adducts in human hemoglobin: structures and mechanisms of formation, Chemical research in toxicology **34**, 1769 (2021).

[62] E. Gianquinto, I. Moscetti, O. De Bei, B. Campanini, M. Marchetti, F. J. Luque, S. Cannistraro, L. Ronda, A. R. Bizzarri, F. Spyrakis, *et al.*, Interaction of human hemoglobin and semi-hemoglobins with the staphylococcus aureus hemophore isdb: a kinetic and mechanistic insight, Scientific reports **9**, 18629 (2019).

[63] S. Dajnowicz, S. Seaver, B. L. Hanson, S. Z. Fisher, P. Langan, A. Y. Kovalevsky, and T. C. Mueser, Visualizing the bohr effect in hemoglobin: neutron structure of equine cyanomethemoglobin in the r state and comparison with human deoxyhemoglobin in the t state, Acta Crystallographica Section D **72**, 892 (2016).

[64] A. Kovalevsky, T. Chatake, N. Shibayama, S.-Y. Park, T. Ishikawa, M. Mustyakimov, S. Z. Fisher, P. Langan, and Y. Morimoto, Protonation states of histidine and other key residues in deoxy normal human adult hemoglobin by neutron protein crystallography, Acta Crystallographica Section D: Biological Crystallography **66**, 1144 (2010).

[65] H.-J. C. Chen, Y.-C. Chen, C.-F. Hsiao, and P.-F. Chen, Mass spectrometric analysis of glyoxal and methylglyoxal-induced modifications in human hemoglobin from poorly controlled type 2 diabetes mellitus patients, Chemical Research in Toxicology **28**, 2377 (2015).

[66] A. M. Stadler, C. Garvey, A. Bocahut, S. Sacquin-Mora, I. Digel, G. Schneider, F. Natali, G. Artmann, and G. Zaccai, Thermal fluctuations of haemoglobin from different species: adaptation to temperature via conformational dynamics, Journal of the Royal Society Interface **9**, 2845 (2012).

[67] S. K. Mahato, A. K. Timiric, M. Mahato, A. Maityb, and Y. K. Waliaa, Studies on the interaction and complexation of pyrrole compound with hemoglobin, J. Indian Chem. Soc **94**, 949 (2017).

[68] F. C. Wireko and D. J. Abraham, The crystal state binding of dithionite to deoxy-hemoglobin, Protein Engineering, Design and Selection **5**, 3 (1992).

[69] S. Sottini, S. Abbruzzetti, F. Spyrakis, S. Bettati, L. Ronda, A. Mozzarelli, and C. Viappiani, Geminate rebinding in R-state hemoglobin: kinetic and computational evidence for multiple hydrophobic pockets, Journal of the American Chemical Society **127**, 17427 (2005).

[70] F. Spyrakis, S. Bruno, A. Bidon-Chanal, F. J. Luque, S. Abbruzzetti, C. Viappiani, P. Dominici, and A. Mozzarelli, Oxygen binding to Arabidopsis thaliana AHb2 nonsymbiotic hemoglobin: evidence for a role in oxygen transport, IUBMB life **63**, 355 (2011).

[71] C. Dalvit and P. E. Wright, Assignment of resonances in

the h^1 nuclear magnetic resonance spectrum of the carbon monoxide complex of human hemoglobin α-chains, Journal of Molecular Biology **194**, 329 (1987).

[72] S. Krzywda, G. N. Murshudov, A. M. Brzozowski, M. Jaskolski, E. E. Scott, S. A. Klizas, Q. H. Gibson, J. S. Olson, and A. J. Wilkinson, Stabilizing bound o2 in myoglobin by valine68 (e11) to asparagine substitution, Biochemistry **37**, 15896 (1998).

[73] J. Vojtechovskÿ, K. Chu, J. Berendzen, R. M. Sweet, and I. Schlichting, Crystal structures of myoglobin-ligand complexes at near-atomic resolution, Biophysical Journal **77**, 2153 (1999).

[74] D. G. Lambright, S. Balasubramanian, S. M. Decatur, and S. G. Boxer, Anatomy and dynamics of a ligand-binding pathway in myoglobin: the roles of residues 45, 60, 64, and 68, Biochemistry **33**, 5518 (1994).

[75] Y. Seno and N. Gõ, Deoxymyoglobin studied by the conformational normal mode analysis: II. The conformational change upon oxygenation, Journal of Molecular Biology **216**, 111 (1990).

[76] J. C. Fromme, A. Banerjee, S. J. Huang, and G. L. Verdine, Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase, Nature **427**, 652 (2004).

[77] S. N. Irizarry, E. Kutluay, G. Drews, S. J. Hart, and L. Heginbotham, Opening the kcsa k+ channel: tryptophan scanning and complementation analysis lead to mutants with altered gating, Biochemistry **41**, 13653 (2002).

[78] U.-M. Ohndorf and R. MacKinnon, Construction of a cyclic nucleotide-gated kcsa k+ channel, Journal of Molecular Biology **350**, 857 (2005).