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. Well-known visualization software allow the computation of conformational change of a particular residue across two different states of a protein. Our formulation takes into account the conformational change of any residue with respect to itself as well as with every other residue. Analyses of twenty eight distinct pairs of proteins from ten distinct classes shows that PRI identifies residues displaying conformational changes bearing influence with respect to itself and all other residues. 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 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 [1]. Generally, a protein's structure is rather flexible, and can undergo several types of structural changes during a biological process [2]. This adaptable nature of a protein's structure is responsible for diverse functional outcomes [3–6]. These structural changes of a protein may be either configurational or conformational [2, 7, 8]. A configurational change in structure might happen due to mutations or other biological processes [9–11]. In this case, the configuration of the amino acid sequence changes [12, 13]. Proteins may also undergo conformational changes, wherein the bonds between the atoms of amino acid sequences remain unchanged [3, 5, 14]. In this case, the dihedral angles of the main chain, and the orientation of side chains change. Conformational changes may occur due to stimuli such as binding of a ligand [15, 16], and environmental changes, such as either mechanical or chemical changes [17]. The change in environment might include the change in solvent composition, ionic strength, pH or temperature [14]. Apart from these, mutations [18], protein folding and misfolding might also trigger conformational fluctuations in a protein [19–21]. Herein, we formulate an *ab initio* information-theoretic approach to study the consequences of residue-level conformational changes and how these shape the function of a protein.

Various experimental techniques such as X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy, cryo-electron microscopy (cryo-EM) are available to understand the three dimensional structure of a protein and the conformational changes that it undergoes in response to stimuli. Computationally, molecular dynamics (MD) simulation provides robust insights into conformational changes of a protein [22, 23]. MD simulations deal with the varied dynamical and conformational structures of a protein, to predict associated functional properties. However, MD simulations are computationally expensive. It is highly challenging to capture delayed conformational changes on timescales outside the scope of MD simulations [24]. 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 multiple network models, such as the Gaussian network model (GNM), elastic network model (ELM), multiscale state space network (MSSN), protein contact network (PCN) or residue interaction network (RIG) *etc.* [25–27]. GNM deals with the dynamic correlations between various conformational structures of a protein [28]. ELM and MSSN enable us to identify residues for conformational transitions. PCN also helps a great deal in studying various conformational states of a protein [29]. PCN finds its application in unraveling details of varied phenomena like protein folding [30, 31], allostery [32, 33], dimer stability [34], residue fluctuation, dynamic structure fluctuation, etc [35]. However, the construction of these network models is non-trivially affected by the choice of the cutoff [36]. Further, these network models often neglect the nature of interactions like van der Waals interactions or hydrogen bonding, and sometimes overlook the complex

* devrani.dbs@presiuniv.ac.in

† soumen@jcbose.ac.in

aspects of the structure of a protein.

The information-theoretic approach proposed herein is plagued neither by computationally expensive demands nor by the vagaries of cutoff distances. Consider any two given states of the same protein, whose experimentally determined structures are available in the Protein Data Bank (PDB). These structures can represent the protein: (a) in two distinct signaling states, (b) before and after ligand binding, or, (c) in two different conformations under various other stimuli. Residues in a protein change their position and/or conformation in response to external stimuli. 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. The first type of change is measured by sequentially aligning the two different structures of the same protein on top of each other and computing the root mean squared deviation (RMSD) value from the residues. The alignment displaying the lowest RMSD value is usually the structure of choice. Common visualisation software enable us to visualise such alignments. 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, and unfolding [20], protein-protein interactions [37, 38] and protein-ligand interactions [39]. Various methods have been introduced to calculate the entropy of proteins [40]. Among these, configuration entropy is used to study interactions and functions of proteins [39, 41]. Configuration entropy also finds application in MD simulations [42, 43] 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 [20, 44]. There also exist other forms of distance-based entropy. However, for example in PCN, the mathematical expression of entropy is influenced by the basic criteria of PCN construction. The interaction between residues is effectively a step function and the choice of cutoff dictates the interaction pattern. This in turn significantly influences the calculation of entropy.

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 crystal 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 [45, 46]. These interactions are mainly responsible for the conformation of side chains of residues of a protein [46]. On the other hand, inter-state entropy aims to distinguish between the entropy of a residue in a given pair of states. Our formulation of a distance-dependent form of entropy is inspired by the extremely well-known Lennard-Jones (LJ) potential [47]. LJ potential is very efficient in capturing the nature of interatomic interactions [48, 49]. In biology, LJ potential finds its application in soft-docking approach [50], molecular dynamics simulations [51], and many others areas [52, 53].

We start with experimentally determined macromolecular structures from PDB. Using the formulation introduced herein, we can compute the *"intra-state entropy"*, *"inter-state entropy"*, and thence the *protein residue information* (PRI) of every residue in the protein. Each of these quantities is rooted in the conformational changes of the protein molecule. By analysing twenty eight distinct pairs of proteins drawn across ten different classes, we discuss extensively about how each of the residues possessing high PRI essay protein function.

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 only deal with crystal structures of a protein in two different states. Therefore, this technique is rather inexpensive computationally. Our findings agree quite well with known experimental evidence in literature.

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}.

The well-known Lenard-Jones (LJ) potential between two interacting entities separated by a distance, r, is $V_{LJ}(r) = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^{6}]$. Here, ϵ and σ are con-

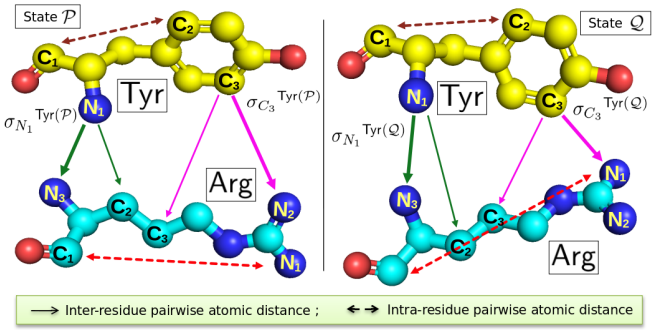

FIG. 1. Illustrative example of two residues of a protein, tyrosine, Tyr, (highlighted in yellow) and arginine, Arg, (highlighted in cyan), undergoing conformational change across the states, $\mathcal{P}$ and $\mathcal{Q}$. Here, we observe that the intra-residue pairwise atomic distances between C_1 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 ${\sigma_{N_1}}^{\text{Tyr}(\mathcal{P})}$ and ${\sigma_{C_3}}^{\text{Tyr}(\mathcal{P})}$ respectively. Similarly, ${\sigma_{N_1}}^{\text{Tyr}(\mathcal{Q})}$ and ${\sigma_{C_3}}^{\text{Tyr}(\mathcal{Q})}$ can be determined analogously for $\mathcal{Q}$. ${\sigma_{N_1}}^{\text{Tyr}(\mathcal{P})} \neq {\sigma_{N_1}}^{\text{Tyr}(\mathcal{Q})}$ and ${\sigma_{C_3}}^{\text{Tyr}(\mathcal{P})} \neq {\sigma_{C_3}}^{\text{Tyr}(\mathcal{Q})}$. Generally, ${\sigma_i}^{\text{k}(\mathcal{P})} \neq {\sigma_i}^{\text{k}(\mathcal{Q})}$, for the ith atom of the kth residue of any protein.

stants. ϵ signifies the strength of interaction and σ represents the distance at which $V_{LJ}(r) = 0$. Obviously, $V_{LJ}(r) > 0$, as two atoms would repel each other when $r < \sigma$. However, for $r > \sigma$, the interaction would be attractive, implying $V_{LJ}(r) < 0$.

Inspired by this well-known formulation, herein we propose, $\omega_{ij}^{kl} = [(\frac{\sigma_i^k}{d_{ij}^{kl}})^6 - (\frac{\sigma_i^k}{d_{ij}^{kl}})^{12}]$, 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. 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 Section III. 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.

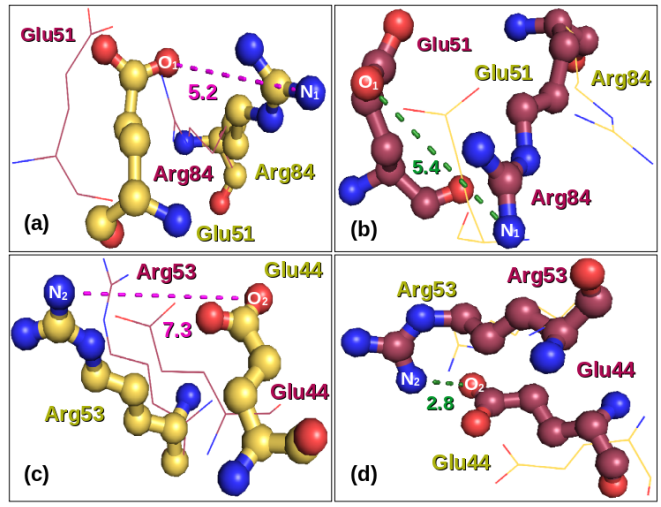

FIG. 2. 1SWE [(a) and (c)] and 1SWA [(b) and (d)] represent the holo (highlighted in golden) and apo (highlighted in maroon) states of streptavidin respectively. The conformation of Glu51 and Arg84 is depicted in (a) and (b), while that of Glu44 and Arg53 is depicted in (c) and (d). The numbers below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

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, $\mathcal{S}(\mathcal{X}) = \sum_{x_i \in \mathcal{X}} -p(x_i) \log_2 p(x_i)$ [54]. 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 crystal 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 including the kth residue. Therefore, $p_i^k = \frac{\sum_{(j,l)} \omega_{ij}^{kl}(d_{ij}^{kl})^2}{\sum_{i=1}^{N_k}\sum_{(j,l)} \omega_{ij}^{kl}(d_{ij}^{kl})^2}$, where, $l \in \{\mathcal{R}\}\backslash\{k\}$ and $j \in \{\mathcal{A}\}\backslash\{i\}$. 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 Supplementary Information (SI).

The corresponding intra-state entropy of the kth residue equals $-\sum_{i=1}^{N_k} p_i^k \log_2(p_i^k)$. The normalised intra-state entropy of the kth residue, $\mathcal{S}_k = -\sum_{i=1}^{N_k} p_i^k \log_2(p_i^k)/\log_2 N_k$, where, N_k, is the total number of atoms present in the kth residue of the protein.

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 crystal 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, $q_i^k = \frac{\sum_{(j,l)}(\Delta_{ij}^{kl})^2}{\sum_{i=1}^{N_k}\sum_{(j,l)}(\Delta_{ij}^{kl})^2}$, for the determination of inter-state entropy. Thence, we can easily compute the normalised inter-state entropy of the kth residue, $\mathcal{J}_k^{(\mathcal{P},\mathcal{Q})} = -\sum_{i=1}^{N_k} q_i^k \log_2(q_i^k)/\log_2 N_k$, 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* $\forall\{\mathcal{P},\mathcal{Q}\}, \mathcal{P} \neq \mathcal{Q}$.

D. Protein Residue Information

Protein residue information (PRI) of residue, k, across two different states, $\mathcal{P}$ and $\mathcal{Q}$, is expressed as, $\mathcal{M}_k^{(\mathcal{P},\mathcal{Q})} = \mathcal{S}_K^{(\mathcal{P})} + \mathcal{S}_k^{(\mathcal{Q})} - \mathcal{J}_k^{(\mathcal{P},\mathcal{Q})}$. 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 us 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, there is a prominent difference between MI and PRI. While, reports of negative MI in quantum information and entanglement exist [55–58], it is well-known to be positive in classical physics. As detailed in the supplementary information (SI), we observe that $\mathcal{M}_k$ can be negative on occasion. This is not unexpected owing to the difference in the physical origin and functional dependence of intra-state and inter-state entropy. As discussed above, $\mathcal{S}_k$ depends on the Euclidean distance between pairs of atoms and is calculated *for any given state*. On the other hand, $\mathcal{J}_k$ depends on the *difference* in Euclidean distance between pairs of atoms and is calculated *across a pair of states.*

A prime motivation of the present work is to identify or predict various residues of a protein with functional importance. Towards that end, we will test whether a residue with higher PRI bears signatures of more "information" associated with conformational changes across two distinct states.

III. RESULTS AND DISCUSSION

We analyse twenty eight distinct pairs of proteins from ten different classes, where each protein posseses a distinct structure in PDB. Each pair possesses two distinct states, $\mathcal{P}$ and $\mathcal{Q}$. Table I displays all residues possessing $\mathcal{M}_k^{(\mathcal{P},\mathcal{Q})} \geq \langle\mathcal{M}\rangle + (1.5 \times S_{\mathcal{M}})$. $\mathcal{M}_k^{(\mathcal{P},\mathcal{Q})}$ 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 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 metaactive 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 (apo-CheY). While the number and nature of identified amino acids differ in different combinations for obvious reasons, we find an interesting trend. Tyr106 is found to be the most important residue and is present in all ten combinations. Indeed, Tyr106 is known to play a very crucial role in the activation of CheY [59–61]. It stabilizes the activated configuration of $\beta4 - \alpha4$ loop and is present in most of the identified allosteric paths [62]. Arg22 is found six times, while Met63 and Lys119 are found five times. Phe14 and Glu89 are found four times. We find Lys7 and Arg19 thrice, while Lys26, Ser56, Lys92 and Lys126 are found twice. Glu27, Glu34, Trp58, Ile95, Phe124 and Met129 are found in one combination. Further, Met63, Lys119 and Arg22 appear whenever 1JBE is considered. Glu89 appears in all comparisons that contain 2B1J, except one. The perturbation propagation algorithm [62] identifies Asp57→ Thr87→ Tyr106→ ligand as the most important allosteric pathway in CheY.

Class of protein-pair	Pairs of PDB structure 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
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
NoIR (Transcription)	4OMZ, 4ON0	Lys12A, Gln61A, Lys65A, Arg75A, Tyr82A, Lys40B, Gln56B, Gln61B, Arg67B, Arg75B
TBX5 (Transcription)	2X6U, 2X6V	Phe57A, Arg82A, Tyr100A, Lys115A, Glu128A, Arg150A, Gln156A, Lys157A, His170A, Cys202A, 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	Lys11A, Trp14A, Tyr24A, Arg31A, Lys60A, Thr67A, Val70A, His72A, Leu83A, Arg141A, Glu22B, Tyr35B, Thr38B, His117B, His146B
	2DN2, 2DN3	Lys16A, Tyr24A, Arg31A, Lys127A, Arg141A, His2B, Trp15B, Glu22B, Tyr35B, Trp37B, Phe71B, 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

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 proteins. 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.

Briefly, phosphorylation at Asp57 (start site) allosterically activates CheY to bind with FLiM (end site). Three distinct molecular events characterise CheY activation: phosphorylation, conformational changes associated with $\beta 4 - \alpha 4$ loop, and rotational restriction of Tyr106 side chain [59]. Phosphorylation of Asp57 causes additional hydrogen bonding interactions with Thr87 preceded by the latter's movement towards phosphorylation site. And the space vacated by Thr87 is occupied by Tyr106. Experiments designate Thr87-Tyr106 coupling as the main reason behind allosteric effects in CheY [62]. Therefore, finding Tyr106 in all ten combinations is no big surprise. Unlike Thr87, the movement of the neighbouring Glu89 appears more distinct towards Tyr106. In FLiM bound state, Glu89 forms a hydrogen bond with Tyr106, and thus stabilizes the activated state of CheY [59]. Indeed, Glu89 lies in several predicted allosteric pathways in CheY [61, 62]. In the presence of mechanical perturbation Glu89 is observed to demonstrate relatively lower thermodynamic response. However, mechanical response is higher [60]. 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. The side chain of Lys119 forms a salt-bridge with Asn16-Film in CheY-N16-FLiM complex [61]. $\beta 4 - \alpha 4$ loop (Ala88-Lys91) is known to be responsible for allostery as well [63]. The adjacent Lys92

undergoes minor differences in respect to bond angles when compared between CheY structures solved at 1.1 and 1.7 resolution [59]. However, in the case of CheY-N16-FLiM complex, hydrogen bonding interactions between Lys92 and Asn16-Flim make the complex stable [61]. Lys109 is an active site residue [60], which behaves as a conformational switch [64]. In fact, Thr87, Ala88, and Lys109 is engaged with phosphoryl group via hydrogen bonding interaction. Phosphorylation-induced conformational changes of $\beta 4 - \alpha 4$ loop along with four residues Glu89, Trp58, Met85, and Tyr106 results in a buried conformation from solvent-exposed form [65]. Met63 changes its rotameric state and comes in closer contact with the adjacent helix. Closer 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 [66]. Mutation of Ile95 to valine enhances the clockwise rotation of the flagellar motor and is correlated with high affinity binding to FLiM. A combination of experimental results identifies the key role of Ile95 in maintaining the surface complementarity for FLiM binding [67]. Direct interaction of Ile95 with FLiM is suggestive of signal transmission through hydrophobic interactions in other response regulators as well. In the Ile95V mutant, besides the Tyr106-Thr87 coupling, Ala88 and Met85 experience significant conformational alterations pertinent to allosteric signaling [65]. 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 crystal structure by influencing the conformation of Asp35 [68]. Further, the synchronised movement of Arg22 and Arg19 abolishes polar contact between the two in FLiM bound state. Arg19 and Glu34 undergo minor conformational changes between CheY95IV and wild-type structures [67]. With subtle conformational differences, Glu34 seems to stabilise Arg22 in the unliganded state. Phe30 and Phe124 sustain conformational changes, which is triggered by Asp57 due to phosphorylation of CheY [64, 69]. The autophospatase activity of CheY is accelerated manifold in the presence of CheZ. However, Lys26 mutation desists the dephosporylation activity of CheZ [70].

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* [71]. In apo-state, the side chain of Ser56 is represented as two distinct rotamers [67]. 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 side-chain conformation shows noticeable difference with the wild-type structure [67]. Further, Glu27 is located in that area of CheY, which helps it in binding to the flagellar switch [64]. 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 [72]. Met129 of *Salmonella enterica* CheY is observed to display high mean field square fluctuation (RMSF) in comparison to *Thermotoga maritima* CheY [73]. Apart from being the terminal residue, the movement of Met129 seems relevant considering the movement of the whole helical segment, upon FLiM binding. Glu93 is located at the solvent-accesible CheA binding surface of CheY [74]. In the activated state, the discrete movement of Glu93 lifts off the steric barrier, thus allowing space for FLiM binding.

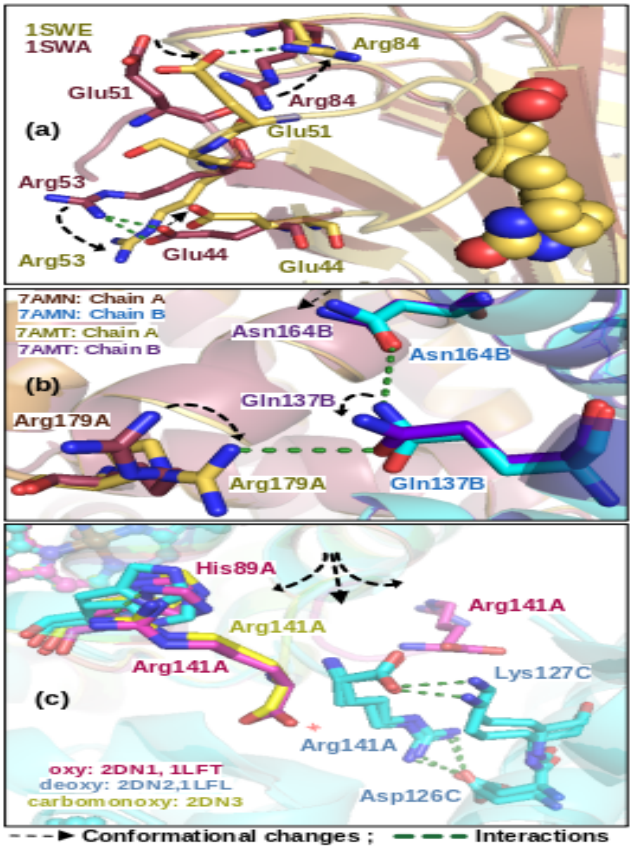

FIG. 3. Formation and destruction of various intra-chain, inter-chain, and inter-subunit interactions in distinct states of: (a) streptavidin, (b) LuxR, and (c) haemoglobin due to conformational changes of residues. (a) Salt-bridge interactions exist between Arg53 and Glu44 in 1SWA. The highlighted contacts 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.

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 [75]. 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. Our study also reveals the same. 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 [76]. 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 tetrameric structure of streptavidin tightens upon binding with the biotin ligand [75]. Experimental observations considering key fluorescence parameters and mutagenesis studies suggest major structural changes in streptavidin tetramer only during the first two biotin binding events. While the third binding event induces structural changes across opposite subunits, changes are insignificant in the fourth, and final binding of the biotin molecule [75]. We, too, observe a difference in the pattern of amino acid residues changing biotin binding. For example, almost the same set of residues is found to be important in biotin binding at subunits B, and C. We do see the involvement of additional residues at subunit C during biotin binding events. Mostly, polar and aromatic residues line the biotin-binding pocket. 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. Extensive polar/hydrogen bonding interactions enhance stability of the complex. We did not identify residues of the surface loop in our study. However, the three adjacent residues Glu51, Ser52 and Arg53, located specifically in subunit C are instrumental in the apo-holo transition. Indeed, a closer view of this portion further reveals significant structural reorientation in this part of the surface loop. 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. Further, the synchronized movement of both Glu51 and Arg84 towards the biotin-binding pocket with the formation of a salt bridge ensures the stable burial of biotin within streptavidin. 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 [77]. Lateral hydrogen bonding interaction between solvent-exposed Tyr22 and Thr20 stabilizes contact regions of higher-order biotinylated streptavidin structures [78]. 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. Inter-residue hydrogen bonding interactions among aromatic amino acids create the necessary 'hydrophobic box' for biotin-binding [76]. Maximum participation of residues from the C subunit followed by that from A, suggests that all subunits participate differently in the biotin-binding process – as evident in fluorescence studies. Differential participation of amino acid residues from four subunits leads to cooperative structural changes and thus allostery in the biotin-binding mechanism.

In the 5A8B-5DKL pair, we investigate the influence of blue light on the conformational dynamics of the signaling aureochrome photoreceptor. Aureochromes 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. Typical of the LOV sensor, aureo1-LOV from *P. tricornatum* (*Pt*Aureo1 LOV) displays the signature five stranded antiparallel β sheet harboring flavin cofactor, 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 [79]. The distribution of residues identified from both the polypeptide chains of dimeric *Pt*Aureo1 LOV is

shown in Figure 1(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 [80]. 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 [81]. 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 [82], 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 [83] and are conserved components of the robust hydrogen bonding network required for LOV photochemistry. Most interestingly, our studies find the residues adjacent to Asn319 and Asn329, namely Tyr320 and Trp328 as detailed in Table 4 and Figure 1(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 [83]. 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.

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 spp.* Therefore, understanding the difference in relative positions of amino acids in LuxR-DNA binding modes during activation (7AMT) and repression (7AMN) sheds insight into the underlying molecular mechanism resulting in such collective behavioral response. LuxR binds with DNA as a homodimer in both states [84]. A quick PDBePISA [85] 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 in activation possesses subtle difference in the repression state, where Asn164A-Gln176 interaction replaces Asn164A-Tyr193B interaction. 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), Gln164A (Gln164B) and Arg179B (Arg179A) triad are exciting. From Figure SI(a), we observe that while transitioning from repression to activation state, Gln137A moves subtly to form hydrogen bonding interactions with Gln164A, moving a little farther from Arg179B in the process.

However, Gln137B breaks the hydrogen bonding interaction by moving farther from Gln164B, 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 [84]. 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 repres-

sion, 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 [86]. 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 inter-state transition. The movement of Ile93B occurs in concert with the movement of Met143B and Phe144B, with both located at the subsequent α_6 helix. 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 during the inter-state transition seem to be the hallmark of LuxR system. 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.

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 theconformational 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 [87] — 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. Interestingly, our recent findings [88] 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 [89]. 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 [87]. 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.

Hemoglobin and myoglobin, which have been studied since the birth of molecular biology are among the most well-characterised biological macromolecules. Herein, we investigate both these systems across different conformational states. 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. The α and β subunits are comprised of 7 and 8 helices respectively. The helices named A to H are joined to each other by non-helical "corners". Heme, nested between E and F helices anchors a ferrous ion via a covalent linkage to the proximal His from F8. The distal site is available for the ligands to bind. Oxygen upon binding to the distal site is stabilised by hydrogen bonding interactions from the neighbouring His (E7). Ligand binding induces series of conformational changes to increase binding affinity. Initiation occurs at the heme binding pocket and extend to the E helix, CD and FG corners,

central water cavity, α_1-β_2 interface as well as other subunits. Experimental studies suggest participation of A helix, F helix and G helix residues as well in modulating Hb allostery. Cooperativity, thus triggered, shifts the allosteric equilibrium from the T-state to R-state [90]. 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. To elaborate, the relaxed state instead of being in the classical R-state is an ensemble of fully liganded states with discrete conformations of quaternary structure. Based on multiple structural parameters like screw-axis rotation/translation, cross-dimer interactions *etc*, these states can be assigned as $R2$, $RR2$, $R3$ and $RR3$. R, $R2$ and $RR2$ represent closed conformations of His(E7)-ligand channel suggesting heme ligand transport. $R3$ and $RR3$ represent open conformations facilitating the release of ligand [90]. Among all the PDBs considered herein, we prioritise two high-resolution structures 2DN1 and 2DN2, representing R and T-states respectively. For comparison with the other ligand bound R-state we consider the carbon monoxide liganded Hb (2DN3) as well. Comparison among tetrameric subunits could have been the best to understand the flow of information from the unliganded to liganded state. However, due to the unavailability of suitable high-resolution structures consisting of two polypeptide chains [α-β] in the asymmetric unit, our calculations are restricted to the flow of information in $\alpha\beta$ subunit of Hb. Residues undergoing relative conformational changes are distributed across Hbs $\alpha\beta$. The respective positions of amino acid residues in the α subunit are Lys11, Trp14, Lys16 in A helix, Tyr24, Arg31 in B helix, Lys60, Thr67, Val70, His72 in D helix, Leu83 in E helix, Arg92 in EF corner and Lys127, Lys139, Arg141 in G helix. The residues at the β subunit are located at Val1, His2 in A helix; Trp15, Glu22, Arg30, Tyr35 in B helix; Trp37 in C helix; Phe71 in E helix; His77 at the EF corner; Lys82 in F helix; His97 at the FG corner; Asn102, Arg104, His117 in G helix and His146 in H helix. Therefore, besides the residues located at A, E, F and G helices, CD/FG corners, we find that amino acid residues located at other helical components (e.g. B, C, D as well as H helices), and corners (e.g. EF) also participate in $T \rightarrow R$ transition. Notably, we find that a large number of amino acids (especially from the D-helix), undergo relative conformational changes while transitioning between unliganded and CO-bound states. Such a trend is not observed during the transition from unliganded to the O_2-bound state. For the identified residues, we find several references in literature. For example, the 'flexible joint' and 'switch' at the $\alpha_1\beta_2$ play a crucial role in $T \rightarrow R$ transition. The aforementioned 'flexible joint' consists of the FG corner, G-helix from chain A (α subunit) and the C-helix, CD corner from chain B (β subunit). The 'switch' region consists of the FG corner, G-helix from chain B and C-helix, CD corner from chain A. Two 'stereochemical differences' have been underlined as the major determinants of $T \rightarrow R$ transition [91]. One prime distinction is the hydrogen bond between Asp94A and Asn102B in the R-state, which is absent in the T-state. Further, Asp99B-Asn97A/Tyr42A hydrogen bonding interactions are present in the T-state but absent in the R-state. In the same region, 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. The Asp94A-Asn102B hydrogen bond connecting the flexible joint and switch region remains intact in the R and $R2$ states. 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 [92]. The C-terminal segments of both A and B are crucial components of Hb tetramer because mutations in this region can potentially disrupt the structure-function of Hb. 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. [90]. 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 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 [90]. 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 [90]. 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 several structural changes with the formation of intra/inter-subunit contacts especially at the $\beta1\beta2$ interface, resulting in a stable T-state structure [90]. 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 [93]. 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. Also, Asn130 is known to participate in the newly formed salt-bridge [94]. Further, mutations in the hinge region (e.g. Tyr35 $\rightarrow$ Arg) of $\alpha 1\beta 2$ perturbs the dimer interface, causing variable affinities towards oxygen [90]. 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 [94]. His146 forms two salt bridges in T-state. However, in the R-state, these salt bridges are removed and His146 gets solvated and disordered [92]. In deoxyhemoglobin, His146 gets delocalised and does not participate in salt-bridge formation [94]. His146 also promotes $T \rightarrow R$ transition. It makes intra-, and intersubunit interactions with $Asp94$, in the case of deoxygenated Hb. His146 is highly disordered in R, $R2$ and $RR3$ structures compared to the T structure. His146 is the critical regulator of the Bohr effect [90]. Tyr24A and Trp15B were observed to take part in nitration of human hemoglobin by $H_2O_2/nitrite$ [95]. The formation of 4-Hydroxybenzyl adduct at Tyr24A in Hb indicates high solvent accessibility [96]. 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 [97]. Lys11A-Ala168B contact is the most conserved interaction at $Isdb^{N1} : Hb$ interface [97]. Lys60A also contributes a hydrogen bond at the $Isdb^{N1} : Hb$ interface [97]. 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 [98], 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 [99]. Another conserved histidine involved in Bohr effect (His117B), however gets deprotonated while transitioning from R to T-state [98]. 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 [100]. Phe71B from the E-helix is located at the frontiers between internal cavities of goblin fold and also known as a mechanically sensitive residue [101]. 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 $\pi - \pi$ interaction between Hb and ligand [102]. 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 [103]. 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 [104]. Location of Val70A at the channel connecting the distal site and solvent suggests its role in ligand migration [105]. 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 [106].

Effective discrimination between oxygen and carbon monoxide by Hb and Mb is a prime factor to aerobic respiration [107]. Therefore, besides Hb we also focus on Mb. We try to explore its residues playing essential role in de-oxy $\rightarrow$ oxy/carbomonoxy transition as well as finding out the basis of varying affinity towards CO and O_2. In the ligand bound conformation, Fe^{2+} in a planar heme surface has a low spin state. Upon the detachment of ligand, Fe^{2+} converts to a high spin state and deviates from planarity by 0.3Å, thereby resulting in a dome-shaped structure. Interestingly, the space filling model of Mb shows no exit/entry pathways for O_2/CO. Therefore, it is certain that conformational changes within Mb are absolutely essential for O_2/CO binding [108]. However, the ligand binding/escape and therefore the opening/closing of Mb molecule is so transient that only time-resolved or high resolution structures to some extent can reveal dynamical features in Mb. Monomeric Mb consists of eight α helices (A to H), connected via flexible corners. 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 [109]. 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. The H64L/V68N mutant shows very high affinity towards oxygen — an additional H-bond from the amide group of Asn (regardless of the nature of amino acid at the 64 position) stabilises the bound Oxygen molecule. While O_2 affinity increases nearly threefold, CO affinity decreases fourfold to sixfold in the V68N mutant. Even though O_2 binding is favoured in the V68N mutant, the Asn side chain sterically inhibits other ligands from approaching the Fe centre through the distal heme binding pocket. 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 bor-

ders the polar heme surface. Lys45 along with His64 and Val68 borders the entry/exit path of diatomic ligand [110]. Further, the mutation of Lys45 by Arg lowers the inner barrier for CO rebinding — as apparent in our findings. Comparison between wild-type/V68N deoxy Mb and V68N Mb-CO highlights His64 undergoing major positional changes in absolute and relative terms. 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 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. High resolution crystal structures, time resolved data and UV Resonance Raman spectroscopy (UVRR) reveal distinct movements of F helix and FG corner away from/towards the heme moiety, upon dissociation of CO. 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. UVRR along with time resolved fluorescence spectroscopy conducted on wild type and Trp mutants reveal notable structural changes between A helix and E helix. This is mostly attributable to the rearrangement of A helix due to the faster motion of E helix, which harbours distal Histidine. While the change in environment around Trp14 is transient, the same around Trp7 is substantial. Incidentally, both these Trp are located at the A helix. Indeed, the neighbouring Gln8 engages with Lys79 at the EF corner and is found to be important for transiting between ligand bound/unbound conformations. Another A helix residue, Lys16 drifts its side chain significantly towards GH corner in O_2 bound state. Calculation of correlation coefficients between "fluctuational motions" of residues as well as heme moiety in deoxy-Mb identified several residues from A, C, E, G and H helix as well as CD, EF, FG and HC corners, possessing strong dynamic interactions with heme [111]. Remarkably, all our identified residues are located at A, C, E, G and H helices as well as CD, EF and FG corners. However, Ref. [111] is only limited to deoxy-Mb. Therefore, not all the residues identified by them feature in our list. 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 also found to be important for unliganded to liganded state. In fact, Lys145 along with Tyr146 are central mediators of dynamic interactions [111] 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.

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 [112], 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. Interestingly, analyses of mutational effects from clinical studies of Halt-Oram reveal that Gly80Arg mutation completely abrogates DNA binding. Unlike Arg81, Gly80 does not interact with DNA. However, our *in silico* mutagenesis studies indicate that Gly80Arg mutation results in severe steric clash with His170, located at the opposite side of Arg81. Therefore, it is clear that the presence of Gly80 is critical not only to achieve DNA binding at the major groove by Arg81, but also to maintain structural stability by accommodating His170 on the opposite side. 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.

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. The catalytic domain of MutY, having strong structural resemblance with Endonuclease III, consists of two structural modules: the 6-helix barrel and another module bearing [4Fe-4S] cluster. MutY interacts with damaged DNA right at the juncture of these two structural modules of the catalytic domain. Additionally, MutY also possesses a C-terminal domain, which is required for OG recognition. Herein, we compare two structures from PDB: 1RRQ and 1RRS. While the former is a co-crystal structure containing MutY bound with damaged DNA bearing 8-OG, the latter is a co-crystal structure of MutY and DNA with 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. The anti-glycosidic bond of 8-OG nucleoside in MutY bound state indicates that this residue abstains from participation in base pairing interaction. However, in absence of MutY, 8-OG attains syn conformation to pair with adenine. *Syn* to *anti* conversion of 8-OG implies a 180° swivelling along the glycosidic bond and an extrusion of adenine from the helix to avoid a potential steric clash. The space thus vacated by adenine can be filled up by the Gln48 side chain. This is succeeded by π-stacking interactions of the amide bearing side chain of Gln48 with the neighbouring $3'$ base as well as hydrogen bonding interactions with the phosphate backbone [113]. 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. 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 CONH2 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. 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). Conformational changes of several amino acid residues are distinctly observed while comparing between the two states. For example, Glu59 side chain rotates by $\approx 90°$ in the opposite direction from Pro29 to be in close contact with Ile55/O. Again, 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. Interestingly, Trp83 does not show any visible conformational changes and remains connected to the backbone carbonyl of Gly87 in both states. 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. 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. Asp160 seems to impart structural stability to the adjacent helix by backbone-side chain hydrogen bonding interactions. The contribution of Trp266 is interesting as its position remains unchanged. However, the bulky non-polar side chain of Trp266 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. Trp352 seems to maintain local hydrophobicity opposite to the DNA

binding face — especially by non-polar interaction with Leu269 and $\pi - \pi$ 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.

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. The selectivity filter sequence TTVGYGD in KcsA channels is conserved in all organisms and precisely regulates the selective yet high-speed movement of K^+ ions across the membrane. The C-type inactivation in KcsA channel not only depends on the conductive $\rightarrow$ constricted conformational changes of the selectivity filter, but also on the residues lining the pore domain [114]. All the residues Arg64, Trp67, Trp68, Ser69 – flank the filter region. 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. A 180° rotation of the indole moiety of Trp67 in 7MHR-7MK6 abolishes hydrogen bonding interaction with the neighboring Asp80 upon gate opening. A strong allosteric coupling of residues in and/or around the filter region ensures the structural integrity of the whole region, and thus K^+ conductivity. Tight interaction between the tip of the helices of S5 and S6 bordering the mouth of the channel is important. Not only does it play a critical role in C-type inactivation in KcsA but also in the coupling between voltage sensor and pore domain in the Drosophila shaker channel. Trp87 is located at the N-terminal end of S6 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. Electron Paramagnetic Resonance (EPR) spectroscopy along with site-directed spin labelling methods reveal significant mobility towards the N-terminal half of the S5 helix and rather remarkable movement towards the C-terminal half of the S6 helix [115]. While Trp26 and Ala28 are present at the N-terminal end of S5, Ala35 showing signs of conformational mobility is located around the middle of S5. Ala35 shows signs of conformational mobility and is located at mid-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 [116]. Leu35 is found in the comparison between 1K4C and 1K4D and directed towards the inter-subunit space. It apparently contributes towards the structural stability of S5 helix. 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. The S6 helix, hinged at Gly99, shows rather drastic conformational changes at these terminal residues – leading to opening/closing of the channel. 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 [117]. 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. The importance of almost all residues identified by us is validated by existing experimental findings from literature. In a small number of cases, the residues identified by us are neighbours to important residues identified experimentally in literature. Their importance is better appreciated in light of the well-known fact that the residues essential in allostery do not always undergo significant conformational changes [118, 119]. During the performance of a specific activity or function, the residues neighbouring an individual residue, rather than the residue itself, rearrange themselves to assist that function. We also predict the likely functional significance of other residues identified by our approach, which can be vali-

dated through experiments.

IV. CONCLUSION

Substantial correlations are usually known to exist between the structure and function of proteins. The position and conformation of residues of a protein may be altered due to diverse factors. Herein, we introduce an *ab initio* information-theoretic approach to study proteins and protein-ligand interactions. The utility of this approach is readily visible as we are largely successful in pinpointing residues undergoing the most significant conformational change — from twenty eight distinct pairs of proteins drawn across ten different classes. As a result of conformational changes between two states of a protein, the position and/or conformation of every residue with respect to its neighbouring and indeed all other residues also get altered. Common visualisation software enable us to visualise such alignments of a residue across two states. The information-theoretic approach presented herein incorporates both these kinds of changes. Both these types of conformational changes are important as they lead to the construction and destruction of various non-covalent interactions, which influence the overall functions of a protein. The identified residues are seen 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.

[1] C. A. Orengo, A. E. Todd, and J. M. Thornton, From protein structure to function, Current Opinion in Structural Biology **9**, 374 (1999).

[2] A. M. Ruvinsky, T. Kirys, A. V. Tuzikov, and I. Vakser, Structure fluctuations and conformational changes in prortein binding, Journal of Bioinformatics and Computational Biology **10**, 1241002 (2012).

[3] L. Orellana, Large-scale conformational changes and protein function: breaking the in silico barrier, Frontiers in Molecular Biosciences **6**, 117 (2019).

[4] S.-R. Tzeng and C. G. Kalodimos, Dynamic activation of an allosteric regulatory protein, Nature **462**, 368 (2009).

[5] B. J. Grant, A. A. Gorfe, and J. A. McCammon, Large conformational changes in proteins: signaling and other functions, Current Opinion in Structural Biology **20**, 142 (2010).

[6] G. Bhabha, J. Lee, D. C. Ekiert, J. Gam, I. A. Wilson, H. J. Dyson, S. J. Benkovic, and P. E. Wright, A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis, Science **332**, 234 (2011).

[7] G. Ungar, E. Aschheim, S. Psychoyos, and D. V. Romano, Reversible changes of protein configuration in stimulated nerve structures, Journal of General Physiology **40**, 635 (1957).

[8] D. Chakravarty, J. Janin, C. H. Robert, and P. Chakrabarti, Changes in protein structure at the interface accompanying complex formation, IUCrJ **2**, 643 (2015).

[9] C. Sotomayor-Vivas, E. Hernändez-Lemus, and R. Dorantes-Gilardi, Linking protein structural and functional change to mutation using amino acid networks, PLOS ONE **17**, 1 (2022).

[10] M. Cusack and W. S. Pierpoint, Similarities between sweet protein thaumatin and a pathogenesis-related protein from tobacco, Phytochemistry **27**, 3817 (1988).

[11] V. M. Chen and P. J. Hogg, Allosteric disulfide bonds in thrombosis and thrombolysis, Journal of Thrombosis and Haemostasis **4**, 2533 (2006).

[12] B. Schmidt, L. Ho, and P. J. Hogg, Allosteric disulfide bonds, Biochemistry **45**, 7429 (2006).

[13] B. Schmidt and P. J. Hogg, Search for allosteric disulfide bonds in nmr structures, BMC Structural Biology **7**, 1 (2007).

[14] J.-H. Ha and S. N. Loh, Protein conformational switches: from nature to design, Chemistry-A European Journal **18**, 7984 (2012).

[15] B. F. Andersen, H. M. Baker, G. E. Morris, S. V. Rumball, and E. N. Baker, Apolactoferrin structure demonstrates ligand-induced conformational change in transferrins, Nature **344**, 784 (1990).

[16] E. Latz, A. Verma, A. Visintin, M. Gong, C. M. Sirois, D. C. Klein, B. G. Monks, C. J. McKnight, M. S. Lamphier, W. P. Duprex, *et al.*, Ligand-induced conformational changes allosterically activate Toll-like receptor 9, Nature immunology **8**, 772 (2007).

[17] D. Molodenskiy, E. Shirshin, T. Tikhonova, A. Gruzinov, G. Peters, and F. Spinozzi, Thermally induced conformational changes and protein-protein interactions of bovine serum albumin in aqueous solution under different ph and ionic strengths as revealed by saxs measurements, Physical Chemistry Chemical Physics **19**, 17143 (2017).

[18] 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).

[19] C. A. Royer, Probing protein folding and conformational transitions with fluorescence, Chemical reviews **106**, 1769 (2006).

[20] A. J. Doig and M. J. Sternberg, Side-chain conformational entropy in protein folding, Protein Science **4**, 2247 (1995).

[21] R. C. Angers, H.-E. Kang, D. Napier, S. Browning, T. Seward, C. Mathiason, A. Balachandran, D. McKenzie, J. Castilla, C. Soto, *et al.*, Prion strain mutation determined by prion protein conformational compatibility and primary structure, Science **328**, 1154 (2010).

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

[23] 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).

[24] 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).

[25] W. Zheng, B. R. Brooks, and G. Hummer, Protein conformational transitions explored by mixed elastic network models, Proteins: Structure, Function, and Bioinformatics **69**, 43 (2007).

[26] C.-B. Li, H. Yang, and T. Komatsuzaki, Multiscale complex network of protein conformational fluctuations in single-molecule time series, Proceedings of the National Academy of Sciences **105**, 536 (2008).

[27] J. G. Su, X. Jin Xu, C. Hua Li, W. Z. Chen, and C. X. Wang, Identification of key residues for protein conformational transition using elastic network model, The Journal of Chemical Physics **135**, 174101 (2011).

[28] I. Bahar, A. Atilgan, and B. Erman, Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential, Folding and Design **2**, 173 (1997).

[29] 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).

[30] G. Bagler and S. Sinha, Assortative mixing in protein contact networks and protein folding kinetics, Bioinformatics **23**, 1760 (2007).

[31] G. Menichetti, P. Fariselli, and D. Remondini, Network measures for protein folding state discrimination, Scientific reports **6** (2016).

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

[33] L. D. Paola, G. Mei, A. D. Venere, and A. Giuliani, Disclosing allostery through protein contact networks, in *Allostery* (Springer, 2021) pp. 7–20.

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

[35] T. J. Lopes, R. Rios, T. Nogueira, and R. F. Mello, Protein residue network analysis reveals fundamental properties of the human coagulation factor VIII, Scientific reports **11** (2021).

[36] S. Sharma, D. Nath, and S. Roy, Allostery and networks, in *Advances in Biology*, Vol. 3 (Nova, New York, USA, 2023) pp. 77–101.

[37] Z. Sun, Y. N. Yan, M. Yang, and J. Z. Zhang, Interaction entropy for protein-protein binding, The Journal of chemical physics **146**, 124124 (2017).

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

[39] 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).

[40] H. Meirovitch, S. Cheluvaraja, and R. P. White, Methods for calculating the entropy and free energy and their application to problems involving protein flexibility and ligand binding, Current Protein and Peptide Science **10**, 229 (2009).

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

[42] S. Chong and S. Ham, Configurational entropy of protein: A combined approach based on molecular simulation and integral-equation theory of liquids, Chemical Physics Letters **504**, 225 (2011).

[43] F. Fogolari, A. Corazza, S. Fortuna, M. A. Soler, B. VanSchouwen, G. Brancolini, S. Corni, G. Melacini, and G. Esposito, Distance-based configurational entropy of proteins from molecular dynamics simulations, PLOS ONE **10**, 1 (2015).

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

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

[46] 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.

[47] J. E. Lennard-Jones, Cohesion, Proceedings of the Physical Society **43**, 461 (1931).

[48] J. B. Adams, Bonding energy models, in *Encyclopedia of Materials: Science and Technology* (Elsevier, Oxford, 2001) pp. 763–767.

[49] T. Iiyama, T. Ohkubo, and K. Kaneko, In situ X-ray diffraction studies on micropore filling, in *Recent Advances in Gas Separation by Microporous Ceramic Membranes*, Membrane Science and Technology, Vol. 6, edited by N. K. Kanellopoulos (Elsevier, 2000) pp. 35–66.

[50] Chapter 4 - best practices for docking-based virtual screening, in *Molecular Docking for Computer-Aided Drug Design*, edited by M. S. Coumar (Academic Press, 2021) pp. 75–98.

[51] L. Kulakova, G. Arampatzis, P. Angelikopoulos, P. Hadjidoukas, C. Papadimitriou, and P. Koumoutsakos, Data driven inference for the repulsive exponent of the lennard-jones potential in molecular dynamics simulations, Scientific reports **7**, 1 (2017).

[52] J. pierre Schermann, Modelling, in *Spectroscopy and Modeling of Biomolecular Building Blocks* (Elsevier, Amsterdam, 2008) pp. 1–57.

[53] F. Stanzione, I. Giangreco, and J. C. Cole, Chapter Four - Use of molecular docking computational tools in drug discovery, Progress in Medicinal Chemistry **60**, 273 (2021).

[54] A. Wehrl, General properties of entropy, Rev. Mod. Phys. **50**, 221 (1978).

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

[56] M. Horodecki, J. Oppenheim, and A. Winter, Quantum state merging and negative information, Communications in Mathematical Physics **269**, 107 (2007).

[57] 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).

[58] F. Caceffo and V. Alba, Negative tripartite information after quantum quenches in integrable systems, arXiv preprint arXiv:2305.10245 (2023).

[59] 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).

[60] 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).

[61] 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).

[62] 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).

[63] 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).

[64] 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).

[65] 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).

[66] 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).

[67] 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).

[68] 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).

[69] 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).

[70] 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).

[71] 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).

[72] 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).

[73] 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).

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

[75] 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).

[76] 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).

[77] 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).

[78] 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).

[79] 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).

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

[81] 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).

[82] 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).

[83] 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).

[84] 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).

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

[86] 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).

[87] 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).

[88] 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).

[89] 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).

[90] 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).

[91] J. Baldwin and C. Chothia, Haemoglobin: The structural changes related to ligand binding and its allosteric mechanism, Journal of Molecular Biology **129**, 175 (1979).

[92] 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).

[93] 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).

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

[95] 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).

[96] 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).

[97] 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).

[98] 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).

[99] 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).

[100] 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).

[101] 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).

[102] 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).

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

[104] 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).

[105] 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).

[106] 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).

[107] 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).

[108] J. Vojtěchovskỳ, 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).

[109] 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).

[110] 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).

[111] 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).

[112] 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).

[113] 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).

[114] A. Rohaim, B. J. Vermeulen, J. Li, F. Kümmerer, F. Napoli, L. Blachowicz, J. Medeiros-Silva, B. Roux, and M. Weingarth, A distinct mechanism of C-type inactivation in the Kv-like KcsA mutant E71V, Nature Communications **13**, 1574 (2022).

[115] E. Perozo, D. Marien, Cortes, and L. G. Cuello, Structural rearrangements underlying K+-channel activation gating, Science **285**, 73 (1999).

[116] S. N. Irizarry, E. Kutluay, G. Drews, S. J. Hart, and L. Heginbotham, Opening the kcsa k+ channel: tryp-

tophan scanning and complementation analysis lead to mutants with altered gating, Biochemistry **41**, 13653 (2002).

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

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

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