Bioinformatics, 31(22), 2015, 3608–3616
doi: 10.1093/bioinformatics/btv429
Advance Access Publication Date: 25 July 2015
Original Paper

Structural bioinformatics

Mapping networks of light–dark transition in LOV photoreceptors

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

[1]Bose Institute, Kolkata 700 009, India and [2]Department of Biological Sciences, Presidency University, Kolkata 700 073, India

*To whom correspondence should be addressed.

Associate Editor: Anna Tramontano

Received on March 5, 2015; revised on July 16, 2015; accepted on July 19, 2015

Abstract

Motivation: In optogenetics, designing modules of long or short signaling state lifetime is necessary for control over precise cellular events. A critical parameter for designing artificial or synthetic photoreceptors is the signaling state lifetime of photosensor modules. Design and engineering of biologically relevant artificial photoreceptors is based on signaling mechanisms characteristic of naturally occurring photoreceptors. Therefore identifying residues important for light–dark transition is a definite first step towards rational design of synthetic photoreceptors. A thorough grasp of detailed mechanisms of photo induced signaling process would be immensely helpful in understanding the behaviour of organisms.

Results: Herein, we introduce the technique of differential networks. We identify key biological interactions, using light-oxygen-voltage domains of all organisms whose dark and light state crystal structures are simultaneously available. Even though structural differences between dark and light states are subtle (other than the covalent bond formation between flavin chromophore and active site Cysteine), our results successfully capture functionally relevant residues and are in complete agreement with experimental findings from literature. Additionally, using sequence-structure alignments, we predict functional significance of interactions found to be important from network perspective yet awaiting experimental validation. Our approach would not only help in minimizing extensive photo-cycle kinetics procedure but is also helpful in providing first-hand information on the fundamentals of photo-adaptation and rational design of synthetic photoreceptors in optogenetics.

Contact: devrani.dbs@presiuniv.ac.in or soumen@jcbose.ac.in

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Optogenetics aims to confer light sensitivity to genetically encoded molecules (Deisseroth *et al.*, 2006), thus controlling multifarious organismal functionalities with great spatiotemporal precision. While the initial motivation might have been the targeting of neuronal pathologies using natural photoreceptors or photosensors, optogenetics has grown beyond the use of naturally existing photoreceptors. With clever accommodation of biologically inspired design of artificial photosensors, optogenetics is now being used extensively to target any significant cellular event. To decipher the synchronization of molecular events among the components of the signaling network, it is essential to understand the signaling states of photosensors in utmost detail. Light sensitive modular proteins usually consist of two parts, photosensor (or input domain, that absorbs light) and effector (or output domain, that executes biological function of that particular protein) and thus enable them to carry out the desired physiological activities such as phototropism, plant morphogenesis, visual perception in animals and so on. Combination of X-ray protein crystallography, ultrafast spectroscopic techniques and mutagenesis studies have identified and validated some of the

major contributing amino acid residues responsible for *light–dark transition*, elaborated later. However, the individuality of photosensors in terms of direction of signal flow, distinct topological arrangement specially in light-oxygen-voltage (LOV) terminals demands further investigation.

LOV domains, first discovered in *A. thaliana* (Huala *et al.*, 1997), have served as candidate sensors for designing light regulated proteins/molecules (Möglich *et al.*, 2009; Strickland *et al.*, 2008). LOV domains are part of Per-Arnt-Sim (PAS) super-family of receptor proteins (Möglich *et al.*, 2010) which control varied biological functions in plants, animals and prokaryotes. LOV photoreceptors absorb blue light by binding to an organic, non-protein component, namely flavin chromophore (FMN or FAD) to initiate photo-cycle which is usually reversible in nature. Photon absorption leads to formation of cysteine-flavin adduct state (Crosson and Moffat, 2001), commonly known as *light state*, resulting in a cascade of signaling processes within the domain. This thermally activated signaling state may sustain itself for timescales of few seconds to few hours and then converts back to the non-signaling state or the so called *dark state*. The most prominent effect of photon-absorption is the formation of cysteine-flavin adduct. Though this is well established for many plant LOVs, a detailed comprehension of the underlying mechanism of the ensuing signal transduction is still in its preliminary stage. For reasons aforementioned, flavin mono nucleotide (FMN) chromophore has not been taken into consideration during network construction and subsequently the conserved cysteine residue, that bonds with FMN chromophore, does not feature in the present analysis.

The goal of this work is to identify important residues in pathway(s) of long-range transmission of structural changes that originate in the chromophore and its immediate environment upon absorption of a photon, by introducing new techniques. Herein, we introduce the technique of differential networks. Knowledge of the aforementioned residues in photoreceptors and their importance is obtained from these differential networks. We successfully identify functionally important residues for the light–dark transition which are in total agreement with experimental findings from previous literature. We also submit predictions on specific functionally important residues which can be experimentally verified. Thus, the methods proposed here should be helpful in mitigating the need of extensive mutagenesis studies (Gleichmann *et al.*, 2013; Zayner *et al.*, 2013). Moreover, precise idea on such participating residues in this transition, would greatly benefit the understanding of molecular mechanisms behind photo adaptation as well as rational design of synthetic photoreceptors for use in Optogenetics.

Network theory (Albert and Barabási, 2002; Newman, 2010) has been applied to many biological systems (Banerjee and Roy, 2015; Filkov *et al.*, 2009; Roy and Filkov, 2009; Roy, 2012) including macromolecules (Amitai *et al.*, 2004; Brinda and Vishveshwara, 2005; Chennubhotla and Bahar, 2006; del Sol *et al.*, 2006; Thibert *et al.*, 2005), and, even towards diagnostics (Banerjee *et al.*, 2015) and biomedical devices (Roy *et al.*, 2015). However, it has never been applied before to study *transition of signaling states in photosensors*. These studies are usually based on structure of the *entire* protein. Not all interactions maybe of equivalent energetic stability. Remarkably, our results demonstrate that at least to a first approximation, overlooking specifics of energetics in interactions (which of course are implicitly included through structural changes during the transition), still leads to good agreement with known experimental findings. For example, for the well researched YtvA and VVD LOV, this agreement is 100 and 88%, respectively.

Both light and dark state crystal structures of any natural or synthetic full-length photoreceptor (not merely the photosensor domain), is simultaneously unavailable till date. It is well-known that upon photon absorption, FMN binding site located in LOV undergoes a major conformational change leading to loss or gain of new interactions in both light and dark states. Therefore, we restrict ourselves to model *only the LOV (sensor) domain, rather than the whole protein* as a network of amino acid residues. It is reasonable to presume that the predicted amino acid residues found to be important for isolated LOV domains might not be too different from full length structures as well. Preservation of coiled coil signaling J_α helix has been evidenced in both the isolated LOV domain of YtvA and in the only crystal structure of full length synthetic photoreceptor structure (Deinsthuber *et al.*, 2013; Gleichmann *et al.*, 2013; Möglich and Moffat, 2007). Crystal structures of core LOV domains show almost no structural difference in dark and light state (Freddolino *et al.*, 2013). However, within the same LOV subfamily that share significant structural homology in core region, there is wide variation in photo cycle kinetics and conformational orientation of the flanking signalling helices. Hence, the sequences and structures are considered individually. To account for these subtle structural differences and to transcend this limitation, herein we introduce the technique of differential networks. We demonstrate that changes arising upon illumination of the LOV sensor domain leading to light–dark transition can be effectively captured by this approach. We show that the combined effect of interactions which are lost or gained during light–dark transition plays an important role in signalling. Detailed analyses of biologically significant interactions among these, yield valuable information.

2 Methods

2.1 Differential network approach

To understand in-depth the light–dark transition, we introduce the approach of *differential network (DN) to specifically characterize interactions unique to each state*. Exposure to light causes formation of new interactions and simultaneous loss of existing interactions within LOV domain. The light (dark) differential network or LDN (DDN) is the network constructed out of edges that are present in the network of a given protein in light (dark) state but absent in dark (light) state. A schematic diagram for construction of DN is provided in Supplementary Figure S1. The differential networks immediately lead us to a much smaller subset of edges which can be scrutinized in far greater detail compared to meticulous studies on the entire parent light or dark network. It is upon deletion of individual edges of DDN or LDN, *in parent light or dark network*; that we are readily able to identify the really important biological interactions.

2.2 Network construction

We conduct detailed analyses on light–dark transition of all LOV photosensors whose dark and light state crystal structures are simultaneously available in Protein Data Bank (PDB). Notably each of the pairs belongs to same space group. They are YtvA from *B. subtilis* (Losi *et al.*, 2005; Möglich and Moffat, 2007) (PDB ids: 2pr6, 2pr5), Vivid (VVD) from the fungus *N. crassa* (Zoltowski *et al.*, 2007) (PDB id: 2pdr, 2pd7), LOV2 domain from *A. sativa* (Oat) (Halavaty and Moffat, 2007) (PDB id: 2v1b, 2v1a), Aureochrome1 (Aureo1) from the photosynthetic marine alga, *V.frigida* (Takahashi *et al.*, 2007) (PDB ids: 3ulf, 3ue6), LOV1 domain of Phot1 from the green alga *C.reinhardtii* (Fedorov *et al.*, 2003) (PDB id: 1n9o, 1n9l)

and LOV2 domain from the phototropin of chimeric fern photoreceptor (Phy3-LOV) (Crosson and Moffat, 2001) (PDB id: 1jnu, 1g28) for light and dark states, respectively. An edge exists between residues i and j, for $(i-j)>2$ (Bhattacharyya *et al.*, 2013) with interaction strength,

$$I=\frac{n_{ij}}{\sqrt{N_iN_j}}, \tag{1}$$

greater than the critical cut-off, I_C. I_C corresponds to the interaction strength at which the size of largest connected component, L, sharply decreases from its initial value at $I=0$. Here n_{ij} is the total number of side chain atom interactions between ith and jth residue within 4.5 Å (Heringa and Argos, 1991). For Glycine, C_α atom is considered (Brinda and Vishveshwara, 2005). N_i and N_j is the total number of heavy atoms in ith and jth residue, respectively (Chennubhotla and Bahar, 2006). For node and edge betweenness, we assign edge weights equal to inverse of interaction strength between amino acid residues. It should be noted that monomer-dimer equilibrium plays a very crucial role in light–dark transition which is discussed extensively in SI. In cases, where the functional form of LOV domain (monomer or dimer) is not known, the network should be constructed for both. Further details about network construction, I_C and space groups are given in Supplementary Information (SI).

2.3 Network metrics

2.3.1 Closeness centrality

C_i, of node i, is the reciprocal of average distance of i to every other node, j, in a network, $\mathcal{G}$, of $\mathcal{N}$ nodes. Thus,

$$C_i=\frac{\mathcal{N}-1}{\sum_j d_{ij}}, \tag{2}$$

where d_{ij} denotes the shortest distance from node i to j.

2.3.2 Betweenness centrality

B_k, of edge (node), k, in $\mathcal{G}$ is a measure of the number of shortest paths passing through k in $\mathcal{G}$. Thus,

$$B_k=\sum_{i\neq k\neq j}\frac{\sigma_{ij}(k)}{\sigma_{ij}}, \tag{3}$$

σ_{ij} being number of shortest paths in $\mathcal{G}$ from edge (node) i to edge (node) j and $\sigma_{ij}(k)$ the number of shortest paths in $\mathcal{G}$ from i to j passing through k.

2.3.3 Eccentricity

e_i, of a node, i, in $\mathcal{G}$ is the maximum of the shortest distance, d_{ij}, between i and every other node j in $\mathcal{G}$. Thus,

$$e_i=\max(d_{ij}) \tag{4}$$

2.3.4 Edge proximity

$\mathcal{P}_k$, of an edge, k, is the inverse of the sum of its shortest distance d_{kl}, to all other edges, l in a network $\mathcal{G}$ with m edges.

$$\mathcal{P}_k=\frac{m-1}{\sum_l d_{kl}} \tag{5}$$

Supplementary Figure S3 in SI schematically introduces the above network metrics. More details are also available elsewhere (Kaur Grewal and Roy, 2015; Roy, 2012)

2.4 Statistical significance factor for k'th edge, $\mathcal{Z}_k$

Change in (a) betweenness of ith node (edge) from average node (edge) betweenness, or, (b) closeness of ith node from average node closeness, or, (c) edge proximity of ith edge from average edge proximity in parent light (dark) state; upon removal of qth edge from parent light (dark) network [or equivalently kth edge of LDN (DDN)],

$$\triangle\mathcal{M}_k^i=\frac{|\mathcal{M}^{\text{parent},i}-\mathcal{M}_q^{\text{remaining},i}|}{\langle\mathcal{M}\rangle} \tag{6}$$

$\mathcal{M}$ represents metric under consideration, i.e. node betweenness, node closeness, edge proximity or edge betweenness. $\mathcal{M}^{\text{parent},i}$ and $\mathcal{M}_q^{\text{remaining},i}$ denote value of $\mathcal{M}$ for ith node or edge of the parent light (dark) network before and after removal, respectively. Thus, $\triangle\mathcal{M}_k^i$ is the change in value of $\mathcal{M}$ for ith residue or edge upon removal of the qth edge of the parent light (dark) network. $\langle\mathcal{M}\rangle$ is the mean value of $\mathcal{M}$ in parent light or dark state averaged over all residues or edges.

Thus, overall effect on $\mathcal{M}$ due to every edge removal is determined as,

$$\triangle\mathcal{M}_k=\frac{1}{(\mathcal{N}-a)}\sum_{i'}\triangle\mathcal{M}_k^{i'} \tag{7}$$

i' denotes number of edges or nodes (other than incident nodes). Thus, $a=0$ for edge betweenness or edge proximity and $a=2$ for node betweenness or node closeness. Finally, statistical significance $\mathcal{Z}_k$ of kth edge of LDN (DDN), or, equivalently the qth edge in the parent light (dark) network undergoing actual removal, is determined by its Z-score,

$$\mathcal{Z}_k=\frac{\triangle\mathcal{M}_k-\langle\triangle\mathcal{M}_k\rangle}{\sigma_{\Delta\mathcal{M}_k}}, \tag{8}$$

where, $\langle\triangle\mathcal{M}_k\rangle$ is the mean effect arising from the individual deletion of every edge of the LDN (DDN) in parent light (dark) network (Eq. 7).

$\sigma_{\Delta\mathcal{M}_k}$ is the associated standard deviation. The kth edge of LDN or DDN is termed as statistically significant edge or key interaction and the corresponding incident nodes are termed as key residues, if $\mathcal{Z}_k\geq 1$. Supplementary Figure S4 in SI clarifies details of $\triangle\mathcal{M}_k$ and statistical significance in a toy network.

3 Results

3.1 Network topology and role of interactions lost or gained during transition

We analyse distribution profiles of different network metrics for the LOV domain, before and after illumination. In Figure 1, the distribution profile for eccentricity implies a common rearrangement pattern among residues in LOV domain of Phy3-LOV and Aureo1-LOV, upon illumination; by and large consistent for all six LOVs as seen in Supplementary Figure S5 of SI. These seem to bear signatures of the effectual strategy employed in presence of light for better signal transduction in LOV. Illumination seems to bring most residues closer to each other and facilitates signaling in the light state. Edge betweenness profiles also show a similar behavior for all LOVs with small number of edges showing very high betweenness values in dark state compared to light state. This long tail in dark state signifies signal flow is primarily channelized through a limited number of edges whereas it is more democratic in the light state. Quantitatively, the standard deviation of both distributions is *always lesser* in light state for all LOV domains studied here as seen in

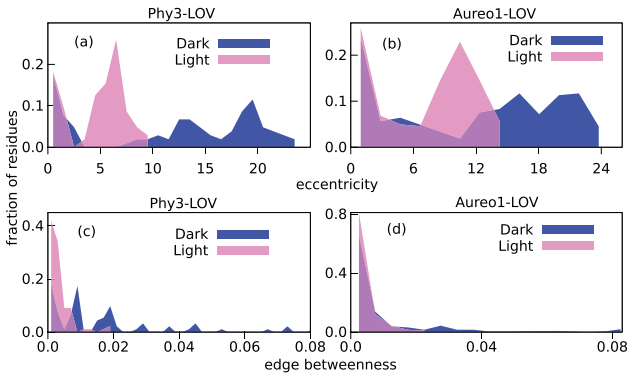

Fig. 1. Distribution of eccentricity (**a** and **b**) and edge betweenness (**c** and **d**) for LOV domains of Phy3 and Aureo1. (a) and (b) Illumination seems to bring most residues closer to each other and facilitates signaling in the light state. Shift of eccentricity distribution towards left in light state shows decrease in shortest communication distance among various residues. (c) and (d) Edge betweenness distribution profile in light state indicates higher participation of most edges in signal flow and lesser 'bottlenecks". Dark state contains significant number of edges with far higher edge betweenness implying that information transfer is primarily channelized via some edges only. Similar shifts are also largely observed for LOVs of other organisms (Color version of this figure is available at *Bioinformatics* online.)

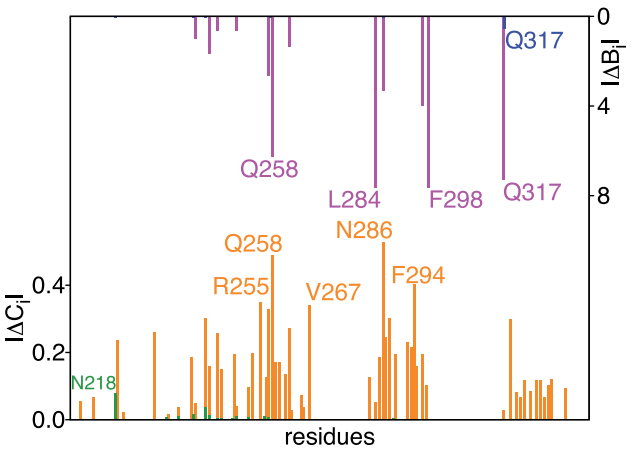

Fig. 2. Effect on closeness and betweenness of residues upon individual deletion of edges (Asn286, Gln258) [significant effect] and (Asn218, Phe238) [trivial effect] for Aureo1-LOV (light state). Effect on closeness (betweenness) for (Asn286, Gln258) is represented by orange (magenta) and for (Asn218, Phe238) by green (blue), respectively. Residues most affected due to any edge deletion are mentioned in appropriate color (Color version of this figure is available at *Bioinformatics* online.)

Supplementary Table S2. Thus, network analysis presents a consistent picture at the level of LOV domain of all organisms, which is achieved without even delving into individual similarities or differences at the molecular level between LOV domains.

It is now imperative to focus on interactions which are *unique to either light or dark state*. Intuitively, one would feel that *absence of interactions* in either the LDN or DDN is quite likely to accommodate a new pattern of information flow in the protein. If an interaction is lost, it is reasonable to infer that its presence would have hindered the signaling process. The nodes associated with such interactions should play a crucial role in signal transmission. Such an important interaction found by us in Aureo1 LOV is (Ile274, Ile270). Ile270 is significant for light-dependent signal transmission (Mitra *et al.*, 2012). Our edge removal strategy successfully captures interactions among residues that are experimentally known to be important either in signal modulation or in functionally important regions.

3.2 Identification of key interactions and residues

All network metrics and statistical measures used here are defined in Section 2. We use node based network centrality metrics like closeness and betweenness and edge based metrics like edge betweenness and edge proximity. The latter has been recently proposed to capture slow-poisoning in networks and seems to be quite important for biological networks as well (Banerjee and Roy, 2015). These global network measures help in comprehending the effect of every single edge deletion on the whole LOV domain. In a given network, a single edge deletion could perhaps significantly affect the betweenness and closeness of nodes *incident on the deleted edge*. However, an average edge deletion usually does not have a major impact on the remaining nodes or edges in the network. Removal of only some specific edges ($\mathcal{Z}_k \geq 1$, see Methods); from the parent light (dark) network, leads to *remarkable change* in network metric values of even distant nodes (as shown in Fig. 2); and/or edges. Thus, edges with $\mathcal{Z}_k \geq 1$ seem to act as a good statistical discriminator and hence we term such edges as *key interactions* and the incident residues as *key residues*. A list of important residues for all six LOV domains, either known experimentally or predicted by us is summarized in Table 1.

As elaborated later, key interactions and key residues found by our method are amply complemented by experimental findings.

3.3 Insights from differential network analysis

We generally find that at least one incident residue of key interactions identified by our method has already been verified experimentally for functional importance. For the well researched YtvA LOV and VVD LOV, this is 100 and 88%, respectively. For Oat LOV and the lesser studied Aureochrome1 LOV, it is 66 and 70%, respectively. The shorter construct for Cr-LOV and near absence of experimental studies on both Cr-LOV and Phy3-LOV result in lower success rate for both of these domains. Inspite of this, nearly two-thirds of the key interactions identified for all LOVs studied here tally with previous experimental findings. Additionally, structure-sequence alignment analysis substantiates the role of key residues identified for all LOVs that await experimental validation (Supplementary Fig. S12). The residues predicted to be important in light–dark transition are mostly localized around FMN binding pocket, signaling helices (A'_α and J_α) and near interface regions as shown in Fig. 3.

3.3.1 YtvA LOV domain (YtvA LOV) from *B. subtilis*

YtvA is responsible for regulating general stress activity in *B. subtilis* in response to blue light. The effector domain linked to core YtvA LOV is an anti-σ factor antagonist (STAS) domain via some linker region (residues 127–147) (Möglich and Moffat, 2007). The J_α helix of YtvA LOV is known to play a significant role in signal transduction. The LDN and DDN identify a set of 22 and 21 edges belonging uniquely to light and dark state, respectively. Our network based approach identifies interactions mostly surrounding the FMN region and few belonging to the dimer region of YtvA LOV. The deletion of (Val90, Leu106) significantly affects the behavior of large number of distant nodes as well as edges. Val90 contributes to the structural stability of YtvA, since its mutation to yet another hydrophobic residue Ile reduces the thermal stability (Song *et al.*, 2013) while Leu106 is involved in dimerization of YtvA LOV (Möglich and Moffat, 2007). Phe46 is known to go through most profound light-induced flipping suggesting flexibility of the domain in this

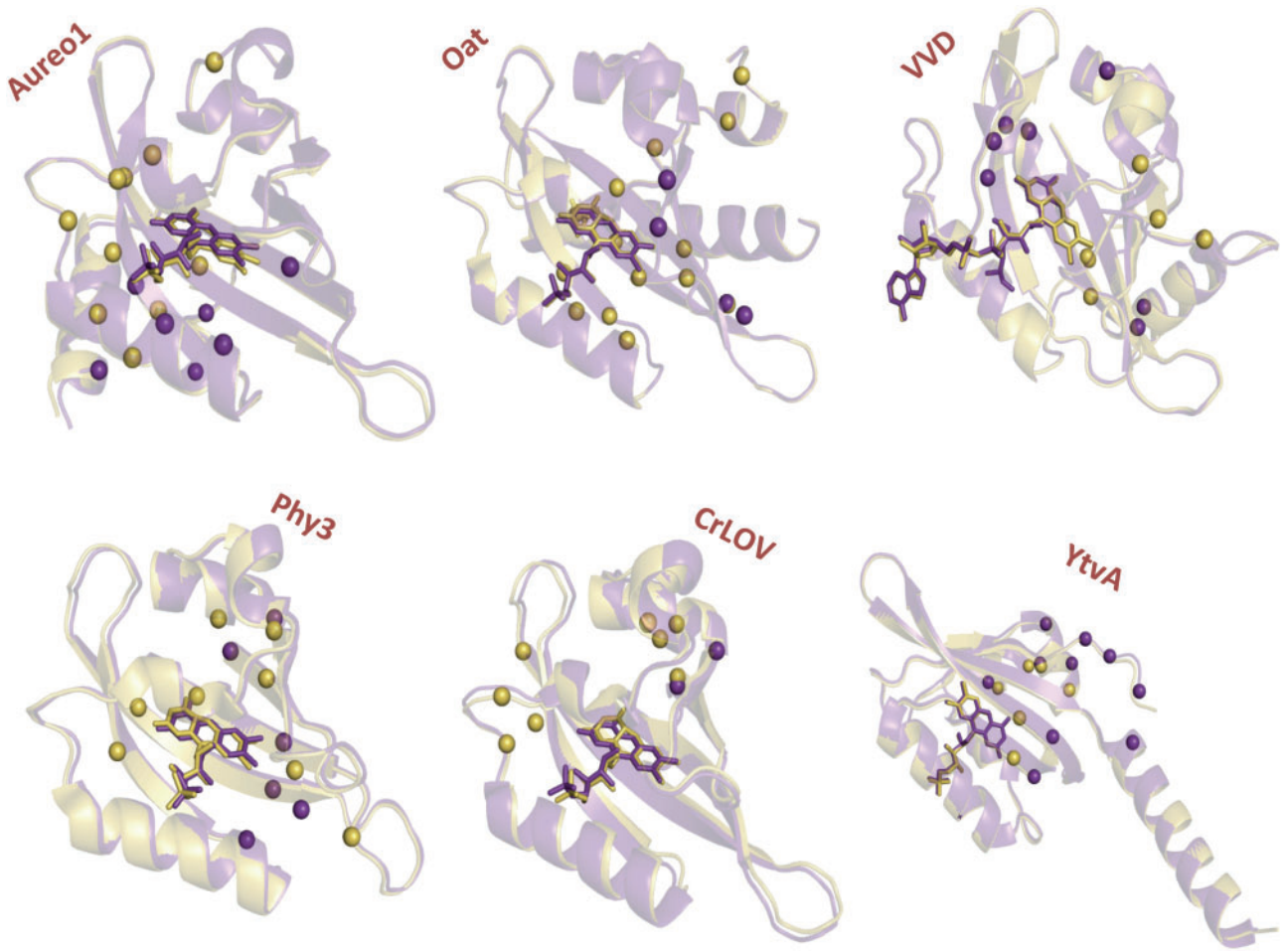

Fig. 3. Distribution of key interacting residues in LOV domain structures of Aureochrome1, Oat, VVD, Phy3, Cr and YtvA The dark and light structures are superimposed and only one protein chain is considered for simplicity, even though some are dimers. Interactions found to be important in LDN (DDN) are shown in brown (purple). Important residues for light and dark states and 3*D* representation of LOVs before and after photon absorption are shown in Supplementary Tables S3–S8 and Supplementary Figures S6–S11, respectively (Color version of this figure is available at *Bioinformatics* online.)

Table 1. Residue interactions identified as functionally important by network approach

Differential Network	YtvA LOV	VVD LOV	Oat LOV	Aureo1 LOV	Cr LOV	Phy3 LOV
Light Differential Network (LDN)	(**V90**, **L106**) (**F46**, M49) (E53, **E56**)	(**Y40**, V86) (**Y40**, L104) (L64, I166)	(A405, **E409**) (**N414**, **Q513**) (**I470**, **F509**) (**V476**, **L496**) (I466, V478) (**L493**, V529) (M499, **I510**)	(K269, V282) (**Q258**, **N286**) (T262, **L284**) (A266, **L284**) (Y244, **L257**) (**L284**, **F298**)	(T21, G40) (P96, L88) (G62, T65) (V52, F41)	(A946, F950) (**N1008**, V996) (V932, **Q1029**) (V1022, **F1025**)
Dark Differential Network (DDN)	(I92, **N104**) (**D21**, Q48) (**V23**, **Q129**) (R24, **Q44**) (**F46**, I57)	(Y98, **L111**) (**Q112**, **N151**) (**I54**, V168) (**M135**, V147) (**M135**, V149) (V149, **N161**)	(P426, G447) (M499, Q507)	(**I270**, **I274**) (C283, Q332) (P204, L212) (V209, **V314**)	(T21, S38)	(S930, S947) (Q1013, I1026) (D991, V1014)

Functional importance of residues of all colours in **bold** have already been experimentally verified in literature. The agreement between predicted residues and known experimental findings is rather good, except for Cr-LOV and Phy3-LOV, which have been poorly studied. Residues that are colored in: (a) brown reside in FMN/FAD binding pocket, (b) blue are located surrounding A'_{α} (signaling helix, N-terminal region) and, (c) magenta are located surrounding J_{α} (signaling helix, C-terminal region) (Color version of this table is available at *Bioinformatics* online.)

particular region (Möglich and Moffat, 2007). Our network analysis effectively captures this behavior of Phe46 which seems to undergo significant rearrangement by gaining (losing) an edge in the light (dark) state. Further, identification of edges (Phe46, Met49) from LDN and (Phe46, Ile57) from DDN, as key interactions emphasizes the prime role played by Phe46 in light-dependent signal transduction. Another key interaction identified for case of LDN is (Glu53, Glu56). Glu56Gln mutation slows down the photocycle twofold (Losi *et al.*, 2005).

On the other hand, in dark state, the edge (Ile92, Asn104) brings out major changes in closeness of distant residues. This result is notable because Ile92 stabilizes its structure through hydrophobic interactions. Upon light absorption, Ile92 and Leu65 undergo displacement from the pseudo-dyad axis of the YtvA structure (Möglich and Moffat, 2007). It is known that Asn104Asp mutation increases the adduct formation rate (Raffelberg *et al.*, 2011). Further experimental observations show that the light induced signal travels through the interactions between residues Ile126, Thr127 and Gln129 of J_α helix with those from H_β and I_β strands. This behavior is prominently featured in present study with (Gln129, Val23) being a key interactions identified for DDN.

Extensive discussion in full-length YF1 structure (Gleichmann *et al.*, 2013) center on mutants that show either constitutive activity in dark or light state (ON variants) or enhanced activity in presence of blue light (INV mutants). Such candidate residues either directly interact with FMN or can be found in A'_α (essential for LOV dimerization) and J_α helices (important for coiled-coil formation and signal transduction to effector domain). Networks analyses agree well with the experimental findings from YF1/YtvA analysis. Large number of residues e.g. Asp21, Val23, Arg24, Gln44, Gln48 as well as Met49 in the YtvA LOV dimer interface have been found to be important in light–dark transition, from networks perspective. Though the functional importance of poorly conserved Glu53 and well-conserved Ile57 in $C_\alpha{-}D_\alpha$ region cannot be ascertained in YtvA, they most possibly confer structural stability to this region. Mutation at their homologous positions in Oat LOV (Zayner *et al.*, 2013) greatly alters $C_\alpha{-}D_\alpha$ helicity and alteration of Ile to Ala decelerates photo-recovery. Therefore, loss or gain of interactions among amino acid residues, as predicted from this work, involve residues that have mostly been validated by structural or kinetic studies. Results are summarized in Supplementary Table S3.

3.3.2 Vivid (VVD) from fungus Neurospora crassa

Vivid is the only LOV known to be devoid of auxiliary domains. Unlike other LOV domains considered here, it consists of flavin adenine di-nucleotide (FAD) chromophore for its light dependent activity. Photo-adduct formation in VVD induces N-terminal conformational changes (Crosson and Moffat, 2001). Network metrics pinpoint three key interactions: (Tyr40, Val86), (Tyr40, Leu104) and (Leu64, Ile166). One of the major molecular events that drives photoreceptor signaling is dimerization of the photosensor domains. Tyr40 plays an important role in dimerization of VVD under light as any mutation to Tyr40 (except with Trp) has resulted in the loss of dimerization ability of VVD under light (Vaidya *et al.*, 2011). A lot has been discussed about the three residues in the B_β strand (I/L-V/I-Y/F) that form evolutionarily conserved signature β-bulge (Zayner *et al.*, 2013) in PAS/LOV domains. It would be interesting to see the role of Val86 of B_β in VVD LOV. Leu104 is almost fully conserved. Homologous Leu446 in Oat LOV accelerates photocycle speed when mutated to glutamate. No functional studies have yet been carried out against Ile166 and its only interaction partner (both from network theory as well as structural analysis) Leu64, unique to FAD binding LOV domains (VVD and WC-1). Interestingly, Gln497 in Oat LOV (homologue of Ile166) forms part of intense hydrogen bonding network and is responsible for maintaining crucial interactions between core LOV domain and J_α helix that unfolds with photo-induction. Since VVD has no J_α helix, we might conclude that the strategic location of H_β residues (like Ile166) could be essential for the transition as Ile166 maintains its only interaction with N-terminal A'_α through Leu64.

In dark state, DDN leads to identification of 6 key interactions: (Asn151, Gln112), (Tyr98, Leu111), (Ile54, Val168), (Met135, Val147), (Met135, Val149) and (Val149, Asn161). Mutations in Ile54 have proven to be deleterious in terms of protein expression and stability. Not just the stability of A'_α per se, but dark–light transition frequently associates with monomer to dimer formation in which surface residues play notable role. Ile54's partner, partially conserved Val168 or any such neutral, nonpolar residue [similar to the effects of homologous Ile219Met mutation (Nakasako *et al.*, 2008)] might be important to maintain dimer structure. Tyr98 is conserved *in toto* across LOV domains and yet, neither the role of Tyr98 not its homologous residues in other LOVs has been investigated. Thus, it might be worth investigating its precise function through experiments. Val149 is mostly present as Leu in other species. Leu480Ala in Oat LOV accelerates photocycle speed (Zayner *et al.*, 2013) and in Aureo1 LOV crystal structure (Mitra *et al.*, 2012), leu284 flips in synchrony with Phe298 at light state. Met135 influences the chromophore's electronic environment (Zoltowski *et al.*, 2009) and exceptionally alters VVD's light state lifetime when interchanged with neutral, nonpolar residues. Val147 is partially conserved and like its homologue Val90 in YtvA, it possibly contributes in structural stability though that awaits experimental validation. Complete set of key interactions is provided in Supplementary Table S4.

3.3.3 LOV2 domain (Oat LOV) from *A. sativa*

We analyse the LOV2 domain of phototropin1 from *A. sativa*. Phototropin1 has a tri-domain structural arrangement consisting of two LOV domains namely LOV1 and LOV2 and a serine/threonine kinase domain (Halavaty and Moffat, 2007). Oat LOV signals through complete disruption of J_α helix from its core domain (Harper *et al.*, 2003) unlike other LOVs considered here.

One of the most prominent findings of Oat LOV dark–light transition that has been experimentally validated is the interaction (Asn414, Gln513), captured distinctively by edge proximity metric. Mutations on the electronegative N414 and Q513 residues, which are in the vicinity of isoalloxazine ring of FMN, identified a potential water gate and water coordination sites. Oat LOV crystal structures at the dark and light states reveal rotations of Asn414 and Gln513 (Halavaty and Moffat, 2007) to allow water molecules, with potential to decay covalent bond through proton transfer to Cys450, enter FMN binding site facilitating dark state restoration (Zayner and Sosnick, 2014). Mutational analysis of Gln513 confirms its role in recovery reaction kinetics (Nash *et al.*, 2008).

To quantitate the extent of conformational changes during dark–light transition in Oat LOV, alteration in circular dichroism helicity (as a measure of unfolding upon illumination) has been thoroughly compared to light-state lifetime (Zayner *et al.*, 2013). Though changes in helicity through Glu409Pro mutation does not alter photocycle speed, it would be interesting to observe the effect of mutation in its partner residue Ala405. Both Val476 and Leu496 when mutated to Phe, moderately increase helicity, but only the former could increase the speed of photocycle. Phe509Leu significantly

increases helicity but decreases photocycle time (Zayner *et al.*, 2013). Both interactions, (Ile466, Val478) and (Met499, Ile510), cause major change in all four node and edge based metrics. Well conserved Val478 possibly confers structural stability similar to its homologous Val90 in YtvA. Ile466 which is also extremely well conserved exception includes Met135 in VVD, which when mutated to Ile reduced the thermal reversion time in VVD. The homologous residue Ile270 in Aureo1 significantly altered the lifetime in light state. Ile217, homologous to Met499, when mutated to Met behaves very similar to WT Arabidopsis Phot2 (Nakasako *et al.*, 2008). Met499 identified through both LDN and DDN, is in a position to hydrogen bond with Ala543 of J_α helix, thus stabilizing the dark state. While it has been shown that any changes in homologous Leu143 in YtvA impairs signal transduction, it remains to be seen as to what would be the effect of Val529 (of J_α helix) alteration in Oat LOV. Although, its interacting partner, Leu493 when mutated with Ala markedly decreases helicity and photocycle time.

For dark state, key interactions correspond to (Gln507, Met499) and (Pro426, Gly447) (Details are in Supplementary Table S5). We have discussed the importance of water gate as well as water coordination sites that have strong influence on the stability of Cys-FMN covalent bond above. Gln507 is strategically located to H-bond with two water molecules immediately at the entry point of water molecules towards chromophore. Though there is no experimental evidence against Gln507 variation in OAT (or T117 in YtvA), it would be interesting to see the effect on photocycle. Residues in the loop regions that essentially confer structural integrity to the photoreceptor structure seem to be important. Examples include conserved Pro426 and Gly447.

3.3.4 Aureo1 LOV domain (Aureo1 LOV) from *V.frigida*

Aureochromes from *V.frigida* are different from other LOV structures due to their reverse effector–sensor topology, which earns them a special place. Their N-terminal effector domain is linked to a C-terminal LOV domain unlike most other known LOV structures. For LDN, only 2 edges, namely, (Lys269, Val282) and (Gln258, Asn286) alter all node and edge based centralities of distant residue or edges. Both residues Gln258 and Asn286 are known to participate in the stabilization of isoalloxazine ring. Val282 located in β-bulge region; H-bonds with Phe298, which is known to undergo prominent conformational changes during light–dark transition. Further, well-conserved Val282 and neighboring Cys283 are in close contact with residues from F_α helix, the flanking helix of β-strands surrounding the chromophore. The interactions of Leu284 with residues Thr262, Ala266 and Phe298 that are unique to light state, all being identified as key interactions, are in support of the experimental evidence of its light-induced conformational change (Mitra *et al.*, 2012). While Thr262 (located just next to the FMN binding pocket) is mostly conserved, Ala266 is poorly conserved. Since Thr462 in Oat LOV (homologous to Ala266) when mutated to Val accelerates photocycle (Zayner *et al.*, 2013), the presence of Ala in this position should enhance the speed of photocycle in Aureo1, as well. Remarkably, Phe298 when mutated increases the light state lifetime 9 fold (relative to wild type) thus obviously playing a significant role in signal transduction (Mitra *et al.*, 2012). Another two residues that significantly alter the light-state lifetime are Ile231 and Ile270, the latter being identified as a crucial node in DDN. Apart from these edges, another edge influencing the betweenness of distant nodes is (Tyr244, Leu257).

Though the role of fully conserved Tyr244 is not experimentally validated even in other LOV species, its strategic location between C_α and D_α helices enables it to stabilize active site Cys254 through a continued H-bonding chain network (Tyr244 → H2O → Leu257 → Cys254) including a water coordination.

In the dark state, key interactions consist of edges (Ile270, Ile274), (Pro204, Leu212), (Cys283, Gln332) and (Val209, Val314). Interestingly Pro204 is also conserved in Vivid and White Collar proteins, also known to transmit signals through N-terminal effector domains. Val209 and partially conserved Leu212 predicted to be important for N-terminal signaling in Aureo1. Results are summarized in Supplementary Table S6.

3.3.5 LOV1 domain of Phot1 from *C.reinhardtii* (Cr-LOV)

For shorter Cr-LOV construct lacking both N-terminal and C-terminal regions (Halavaty and Moffat, 2007) and very few related experimental studies, residues identified as crucial for dark–light transition lack experimental validation. Therefore, we validate our results mostly through structure-sequence alignment. Interestingly, most of the homologous residues corresponding to key residues of Cr-LOV are found to be important given the mutational analysis of other well-studied photoreceptors like YtvA and Oat LOV. For light state, we have four key interactions: (Thr21, Gly40), (Pro96, Leu88), (Gly62, Thr65) and (Val52, Phe41) while for DDN, our approach identifies a unique key interaction (Thr21, Ser38). Results are summarized in Supplementary Table S7.

H-bonding network around the chromophore has always been found significant for the stabilization of light and dark states. Thr21 features in both LDN and DDN H-bonds with Gln120. In turn, Gln120 directly H-bonds with FMN oxygen atom. The neighboring residue of Thr21 and also its interaction partners Gly40 (in LDN) and Ser38 (in DDN) stabilize C_α helix through H-bond. (Thr21, Gly40) and (Thr21, Ser38) interactions seem to stabilize the scaffold of LOV structural components in the light and dark states, respectively. Both the homologous residues of conserved Phe41 (Phe46 in YtvA and Phe434 in Oat LOV) undergo prominent light-induced structural changes (Möglich and Moffat, 2007) and their variants substantially alter the photocycle speed (Raffelberg *et al.*, 2011; Zayner *et al.*, 2013). Partially conserved Val52 (Ile in other cases) perhaps maintains photocycle stability. Mutation to a smaller hydrophobic group (Ile445Ala) remarkably increases helicity in the light state and decelerates photocycle speed in Oat LOV. Gly62 and Thr65, conserved in all FMN-binding LOV, belong to H-bonding network that stabilizes FMN binding domain (interactions through Leu88-Asn89 and Asn61). Partially conserved Leu88 and Pro96 are components of β bulge. Homologous residues in YtvA are displaced towards the outside, possibly as an exercise of conformational rearrangement upon light activation.

3.3.6 LOV2 domain of phototropin from chimeric fern photoreceptor (Phy3-LOV)

LDN of Phy3-LOV identifies four key interactions. As we observe the dark and light state crystal structures of Phy3-LOV (Crosson and Moffat, 2001, 2002), a clear movement of Gln1029 residue is observed through dark–light transition. In light-state, concomitant to the new bond formation between FMN *N5* and Cys966, H-bond between Gln1029 and Ser930 is broken to facilitate bonding between Gln1029 and Gln1027. Such changes are also evident in the network analysis here, where both Gln1029 and Ser930 are found to be important in LDN and DDN, respectively. The residue homologous to Gln1029 in Oat LOV is Gln513, whose role in dark–light transition has already been discussed in detail. Another homologue, Gln123 in YtvA interacts directly with FMN (Gleichmann *et al.*,

2013). Val932, partner of Gln1029, belongs to the hydrophobic pocket surrounding the dimethyl benzene moiety of FMN (Crosson and Moffat, 2001). Mutations of its homologs, Val416ALa in Oat LOV and Ile74Val in VVD LOV, alter photocycle speed.

Interacting partner of Asn1008, Val996 is mostly Leu at the equivalent position in other LOV domains. Homologs Leu480 in Oat LOV greatly enhances photocycle speed and Leu284 in Aureo1 undergoes concerted movement along with Phe298 after photo-excitation. Another important finding is the (Phe950, Ala946) interaction. Phe950's homologue, Phe46 in YtvA undergoes most significant light induced conformational changes (Möglich and Moffat, 2007). Superimposition of Phy3 light versus dark structure does show movement of Phe950 ring upon illumination. However, the changes are not so prominent as in YtvA structure. It would be interesting to see if Phe950 can alter the photocycle speed akin to Phe46 in YtvA (Raffelberg *et al.*, 2011) or Phe434 in Oat LOV (Zayner *et al.*, 2013). Though there is no apparent evidence against the partially conserved partner of Phe950 (namely Ala946), providing support to neighboring Phe950 through hydrophobic interactions, it might be important for structural integrity. Another such example could be the interaction between Phe1025 (part of FMN binding pocket and β-bulge) and Val1022 in the adjacent loop.

There are three key interactions in Phy3 DDN. While the role of Ser930 is already established in LDN, the function of its partner residue, Ser947 (which is almost fully conserved), is unknown. Straddled between B_β and C_α, Ser947 is a part of the intense H-bonding network in this region, especially with the residues from A_β. Throughout the network analyses conducted here, we find that residues, even those which do not belong to chromophore binding pocket, are featured to be important. And one of the major reasons seems to be their participation in intense H-bonding network, which is an integral part of structural stability. More such residues found significant from Phy3 DDN are Gln1013, Val1014 and Ile1026, located on the chromophore-lining β-scaffold and Asp991. It has been shown that Gln497 (Gln1013 homologue in Oat LOV (Halavaty and Moffat, 2007)); is among the few residues that forms an extensive H-bonding network between core LOV domain and J_α helix, thus essential for photo-induced signaling. Details appear in Supplementary Table S8.

4 Discussion

FMN/FAD based blue-light responsive LOV photoreceptors are present across all kingdoms of life. This indicates their indispensability and diverse function in many crucial biological signaling events. Photon absorption leads to signal propagation accompanied by structural alterations, downstream or upstream, which eventually affects the behavioral pattern of the organism. Since our study is completely dependent on crystal structure, it should be noted that photo-induced dynamics amidst static nature might vary with data collection at different temperature. Further, crystallography artefacts might also mildly affect network construction from different chains, even from one single structure.

Nevertheless, by introducing the notion of differential network, our study successfully identifies important interactions and residues with functional significance, as attested by experiments as well as sequence-structure homology. Thus, it would be helpful in guiding future experiments. Our network based approach is not only useful in understanding photo-adaptability in natural photoreceptors but will be extremely helpful in rational design of novel and complex photoreceptors with interesting light-responsive functions.

Acknowledgements

We thank Prof. Keith Moffat, Prof. KP Das and anonymous reviewers for critical comments on the manuscript and Prof. Saraswathi Vishveshwara and Prof. Gautam Basu for useful discussions. R.K.G. acknowledges financial support from University Grants Commission of India.

Conflict of Interest: none declared.

References

Albert,R. and Barabási,A.L. (2002) Statistical mechanics of complex networks. *Rev. Mod. Phys.*, **74**, 47–97.

Amitai,G. *et al.* (2004) Network analysis of protein structures identifies functional residues. *J. Mol. Biol.*, **344**, 1135–1146.

Banerjee,S.J. *et al.* (2015) Slow poisoning and destruction of networks: edge proximity and its implications for biological and infrastructure networks. *Phys. Rev. E*, **91**, 022807.

Banerjee,S.J. *et al.* (2015) Using complex networks towards information retrieval and diagnostics in multidimensional imaging. arxiv:1506.02602

Bhattacharyya,M. *et al.* (2013) An automated approach to network features of protein structure ensembles. *Protein Sci.*, **22**, 1399–416.

Brinda,K.V. and Vishveshwara,S. (2005) A network representation of protein structures: Implications for protein stability. *Biophys. J.*, **89**, 4159–4170.

Chennubhotla,C. and Bahar,I. (2006) Markov propagation of allosteric effects in biomolecular systems: application to GroEL-GroES. *Mol. Syst. Biol.*, **2**, 36.

Crosson,S. and Moffat,K. (2001) Structure of a flavin-binding plant photoreceptor domain: insights into light-mediated signal transduction. *Proc. Natl Acad. Sci. USA*, **98**, 2995–3000.

Crosson,S. and Moffat,K. (2002) Photoexcited structure of a plant photoreceptor domain reveals a light-driven molecular switch. *Plant Cell*, **14**, 1067–1075.

Deinsthuber,R.P. *et al.* (2013) Full-length structure of a sensor histidine kinase pinpoints coaxial coiled coils as signal transducers and modulators. *Structure*, **21**, 1127–1136.

Deisseroth,K. *et al.* (2006) Next-generation optical technologies for illuminating genetically targeted brain circuits. *J. Neurosci.*, **26**, 10380–1386.

del Sol,A. *et al.* (2006) Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol. Syst. Biol.*, **2**, 2.

Fedorov,R. *et al.* (2003) Crystal structures and molecular mechanism of a light-induced signaling switch: the Phot-LOV1 domain from *C. reinhardtii*. *Biophys. J.*, **84**, 2474–2482.

Filkov,V. *et al.* (2009) Modeling and verifying a broad array of network properties. *Europhys. Lett.*, **86**, 28003.

Freddolino,P.L. *et al.* (2013) Signaling mechanisms of LOV domains: new insights from molecular dynamics studies. *Photochem. Photobiol., Sci.* **12**, 1158–1170.

Gleichmann,T. *et al.* (2013) Charting the signal trajectory in a Light-Oxygen-Voltage photoreceptor by random mutagenesis and covariance Analysis. *J. Biol. Chem.*, **288**, 29345–29355.

Halavaty,A.S. and Moffat,K. (2007) N- and C-Terminal flanking regions modulate light-induced signal transduction in the LOV2 domain of the blue light sensor Phototropin 1 from *Avena sativa*. *Biochemistry*, **46**, 14001–14009.

Harper,S.M. *et al.* (2003) Structural basis of a phototropin light switch. *Science*, **301**, 1541–1544.

Heringa,J. and Argos,P. (1991) Side-chain clusters in protein structures and their role in protein folding. *J. Mol. Biol.*, **220**, 151–171.

Huala,E. *et al.* (1997) Arabidopsis NPH1: a protein kinase with a putative redox-sensing domain. *Science*, **278**, 2120–2123.

Kaur Grewal,R. and Roy,S. (2015) Modeling proteins as residue interaction networks. *Protein Peptide Lett.*, doi: 10.2174/0929866522666150728115552.

Losi,A. *et al.* (2005) Mutational effects on protein structural changes and interdomain interactions in the blue-light sensing LOV protein YtvA. *Photochem. Photobiol.*, **81**, 1145–1152.

Mitra,D. *et al.* (2012) Crystal structures of Aureochrome1 lov suggest new design strategies for optogenetics. *Structure*, **20**, 698–706.

Möglich,A. and Moffat,K. (2007) Structural basis for light-dependent signaling in the dimeric LOV Domain of the photosensor YtvA. *J. Mol. Biol.*, **373**, 112–126.
Möglich,A. *et al*. (2009) Design and signaling mechanism of light-regulated histidine kinases. *J. Mol. Biol.*, **385**, 1433–1444.
Möglich,A. *et al*. (2010) Structure and function of plant photoreceptors. *Ann. Revs. Plant. Biol.*, **61**, 21–47.
Nakasako,M. *et al*. (2008) Structural basis of the LOV1 dimerization of *Arabidopsis* Phototropins 1 and 2. *J. Mol. Biol.*, **381**, 718–733.
Nash,A.I. *et al*. (2008) A conserved glutamine plays a central role in lov domain signal transmission and its duration. *Biochemistry*, **47**, 13842–13849.
Newman,M.E.J. (2010) *Networks: An Introduction*. Oxford University Press, Oxford, UK.
Raffelberg,S. *et al*. (2011) Modulation of the photocycle of a lov domain photoreceptor by the hydrogen-bonding network. *J. Am. Chem. Soc.*, **133**, 5346–5356.
Roy,S. and Filkov,V. (2009) Strong associations between microbe phenotypes and their network architecture. *Phys. Rev. E*, **80**, 040902 (R).
Roy,S. (2012) Systems biology beyond degree, hubs and scale-free networks: the case for multiple metrics in complex networks. *Syst. Synth. Biol.*, **6**, 31–34.
Roy,S. *et al*. (2015) A system and method for analyzing videos of application or function for feature identification of the videos and related application or function. Indian Patent 628/KOL/2015.
Song,X. *et al*. (2013) Engineering a more thermostable blue light photo receptor *Bacillus subtilis* YtvA lov domain by a computer aided rational design method. *PLoS Comput. Biol.*, **9**, e1003129.
Strickland,D. *et al*. (2008) Light-activated DNA binding in a designed allosteric protein. *Proc. Natl Acad. Sci. USA*, **105**, 10709–10714.
Takahashi,F. *et al* (2007) AUREOCHROME, a photoreceptor required for photomorphogenesis in stramenopiles. *Proc. Natl Acad. Sci. USA*, **104**, 19625–19630.
Thibert,B. *et al*. (2005) Improved prediction of critical residues for protein function based on network and phylogenetic analyses. *BMC Bioinf.*, **6**, 213.
Vaidya,A.T. *et al*. (2011) Structure of a lightaActivated LOV protein dimer that regulates transcription. *Sci. Signal.*, **4**, ra50.
Zayner,J.P. *et al*. (2013) Investigating models of protein function and allostery with a widespread Mutational Analysis of a Light-Activated Protein. *Biophys. J.*, **105**, 1027–1036.
Zayner,J.P. and Sosnick,T.R. (2014) Factors that control the chemistry of the LOV domain photocycle. *PLoS One*, **9**, e87074.
Zoltowski,B.D. *et al*. (2007) Conformational switching in the fungal light sensor Vivid. *Science*, **316**, 1054–1057.
Zoltowski,B.D. *et al*. (2007) Mechanism-based tuning of a LOV domain photoreceptor. *Nat. Chem. Biol.*, **5**, 827–834.