Modular Diversity of the BLUF Proteins and Their Potential for the Development of Diverse Optogenetic Tools

Manish Singh Kaushik 1, Ramandeep Sharma 1, Sindhu Kandoth Veetil 1, Sandeep Kumar Srivastava 2,* and Suneel Kateriya 1,*

1 School of Biotechnology, Jawaharlal Nehru University, New Delhi 110067, India; manish13587@gmail.com (M.S.K.); ramandeep31081995@gmail.com (R.S.); kvsindhu@gmail.com (S.K.V.)
2 Department of Biosciences, Manipal University Jaipur, Jaipur 303007, India
* Correspondence: sandeepkumar.srivastava@jaipur.manipal.edu or findsandeep1@gmail.com (S.K.S.); skateriya@jnu.ac.in (S.K.)

Received: 7 April 2019; Accepted: 6 May 2019; Published: 19 September 2019

Abstract: Organisms can respond to varying light conditions using a wide range of sensory photoreceptors. These photoreceptors can be standalone proteins or represent a module in multidomain proteins, where one or more modules sense light as an input signal which is converted into an output response via structural rearrangements in these receptors. The output signals are utilized downstream by effector proteins or multiprotein clusters to modulate their activity, which could further affect specific interactions, gene regulation or enzymatic catalysis. The blue-light using flavin (BLUF) photosensory module is an autonomous unit that is naturally distributed among functionally distinct proteins. In this study, we identified 34 BLUF photoreceptors of prokaryotic and eukaryotic origin from available bioinformatics sequence databases. Interestingly, our analysis shows diverse BLUF-effector arrangements with a functional association that was previously unknown or thought to be rare among the BLUF class of sensory proteins, such as endonucleases, tet repressor family (tetR), regulators of G-protein signaling, GAL4 transcription family and several other previously unidentified effectors, such as RhoGEF, Phosphatidyl-Ethanolamine Binding protein (PBP), ankyrin and leucine-rich repeats. Interaction studies and the indexing of BLUF domains further show the diversity of BLUF-effector combinations. These diverse modular architectures highlight how the organism’s behaviour, cellular processes, and distinct cellular outputs are regulated by integrating BLUF sensing modules in combination with a plethora of diverse signatures. Our analysis highlights the modular diversity of BLUF containing proteins and opens the possibility of creating a rational design of novel functional chimeras using a BLUF architecture with relevant cellular effectors. Thus, the BLUF domain could be a potential candidate for the development of powerful novel optogenetic tools for its application in modulating diverse cell signaling.

Keywords: photoreceptor; BLUF; modular domain; optogenetics

1. Introduction

Microorganisms respond to changing light conditions using an evolved repertoire of photoreceptors that perceive light and execute a light-dependent control of regulatory ‘output’ domains [1]. Blue-light using flavin (BLUF) protein photoreceptors respond to blue light and are often coupled with different effector domains (enzymes or transcriptional regulators) to generate the full range of combinations to regulate photo-adaptive responses [2–5]. However, proteins having the BLUF domain with an extended C-terminus only have also been reported, and their responses were controlled by light-dependent protein-protein interactions [6–9]. Upon illumination, the isoalloxazine
moiety of the BLUF domain associated flavin chromophore (flavin adenine dinucleotide; FAD, or flavin adenine mononucleotide; FMN, or riboflavin; RF) absorbs blue light [10], and undergoes structural rearrangements to modulate the communion between BLUF and the effector domains [11]. Unlike the complex mechanisms of photo-transformation in other photoreceptors, the BLUF domain, upon illumination, mainly shows a hydrogen bond rearrangement around the flavin cofactor, which causes a 10–15 nm red shift in the BLUF absorbance peak [11]. The photo-activation of the BLUF domain is due to the involvement of a conserved glutamine and tyrosine residues [12–14]. The hydrogen bond rearrangement involves a unique ability of the BLUF domain, i.e., photo-induced proton-coupled electron transfer (PCET), which enables them to switch between receptor and signaling states [15–17]. The photo-activation and structural rearrangements around the chromophore of the BLUF domain are transmitted as a signal for the activation of an associated effector domain. The BLUF domains are considered as an attractive model to investigate new paradigms of photo-induced signaling. BLUF domains have a modular architecture; hence, they may be functionally fused to different effector domains, as observed for the modular light oxygen voltage (LOV) photoreceptors [18–25]. Barends and coworkers have characterized a full-length active photoreceptor, BlrP1, from *Klebsiella pneumoniae*, which is composed of BLUF and EAL (Glutamine-Alanine-Leucine) as the sensor/output domain combination, respectively [4]. The EAL domain is a conserved signature motif which hydrolyses cyclic dimeric GMP (c-di-GMP) and is involved in the regulation of motility, biofilm formation, virulence and antibiotic resistance in the bacteria [26–29]. When exposed to light, the BLUF domain from BlrP1 activates the EAL domain via an allosteric communication relayed through conserved domain-domain interfaces [4]. In *Escherichia coli*, YcgF is another photoactivated protein with the BLUF-EAL domain that has been reported [30]. However, unlike other EAL domain proteins, YcgF acts as a transcriptional regulator and controls the YcgF/YcgE pathway, which regulates the synthesis of small regulatory proteins. These small regulatory proteins are involved in the modulation of biofilm functions via the Rcs two-component pathway, necessary for *E. coli* to sustain the adverse environment [30]. In many bacteria, proteins with tandem GGDEF (diguanylate cyclase; DGC)/EAL (phosphodiesterase) domains were also reported to be involved in the c-di-GMP turnover, which modulates a variety of functions ranging from the functional modification of cell surface components, the expression of extracellular signaling molecules, virulence and motility [31,32]. Photoactivated BLUF associated adenylyl cyclase homology domains (CHD) were also investigated and well characterized in several microorganisms, where they are specifically involved in the catalytic conversion of ATP to cyclic AMP (cAMP), which regulates the downstream signal transduction [33–38]. PAS domains (Per/ARNT/Sim) are one of the broadly spread domains involved in sensing variations in light, oxygen, redox potential and the binding of small ligands [39]. The role of PAS domains is diverse, and few reports demonstrated the involvement of PAS in domain chromophore attachment [40], light-regulated protein-protein interactions [41] and in complementary chromatic adaptation (CCA) [42].

Optogenetics is a recently developed molecular tool that combines genetic and optical methods and enables us to modulate specific functions in any cell or tissue using light in a controlled manner [43]. Microbial, algal opsins and natural light regulated ion channels have been reported to be versatile and good actuators for optogenetic applications [44]. However, BLUF domains, due to their small size, solubility, reversibility, temporal precision and diverse association with a wide variety of effectors, could be engineered for the light-dependent modulation of a wide range of cellular signaling [44]. Prominent examples include BLUF containing photoactivated adenylyl cyclase from *Beggiatoa* sp. (bPAC) efficiently activating cyclic-nucleotide-gated ion channels in neurons. The mutagenic variant of bPAC, BlaG, yielded a higher level of light-induced production of cGMP than cAMP [11,45]. In the present study, we have characterized the modular diversity of the BLUF domain coupled proteins that could be valuable in the development of novel synthetic photoswitches and we expand the scope of the optogenetics modulation of novel cellular signaling within a functional expression in the appropriate living system.
2. Materials and Methods

2.1. Database of Sequences used in this Analysis

The BLUF domain encoding protein sequences were retrieved from the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/), and each of them was subjected to a conserved domain search using the Conserved Domain Architecture Retrieval Tool (CDART; https://www.ncbi.nlm.nih.gov/Structure/lexington/ncbi.cgi) [46]. The 34 uncharacterized BLUF domain containing proteins were selected for a further analysis. For each protein, sequences encoding the BLUF domain were selected and aligned for the homology analysis using the BioEdit tool [47]. The Multiple EM for the Motif Elicitation (MEME) suite (http://meme-suite.org) was employed to scan conserved motifs throughout the sequences [48]. The interacting partners for each of the output domains were predicted using the String version 11 [49].

2.2. Phylogenetic Analysis

A phylogenetic analysis involving 34 BLUF sequences from different organisms was performed by employing the Maximum Likelihood method, based on the JTT matrix-based model [50]. Gaps were eliminated from the sequences. The tree with the highest log likelihood (−955.3175) is shown. The values shown with the branches represent the percentage of trees clustering with the associated taxa. The JTT matrix of pairwise distance was subjected to Neighbor-Join and BioNJ algorithms for the construction of the first tree(s), which was then used to select the topology with a higher log likelihood value. Phylogenetic analyses were performed using MEGA6 [51].

2.3. Analysis and Homology Modeling of the BLUF Domain

The annotated sequences were further analyzed for the putative secondary structures and function using Predict Protein (https://www.predictprotein.org/). Based on the secondary structure analysis, a two-dimensional topology was generated using POTTER (http://wlab.ethz.ch/protter/#) [52]. Three-dimensional models of the predicted BLUF sequences and associated effector domains were created using the Phyre2 modeling tool [53], employing an integrated combinatorial approach comprising comparative modeling, threading, and ab initio modeling [54]. All the energy-minimized models of the annotated BLUF domains and BLUF, in combination with the effector domains, were further evaluated for structural errors and the stereochemistry quality, as well as for manual curation. The variety of the models were performed, in terms of bond angles, distances, stereochemical analysis, and vocabulary, by the UCLA Structure Analysis and Verification Server (SAVES) with the PROCHECK and ERRAT programs [55,56]. Finally, the most acceptable models were finalized based on Ramachandran plot analysis and structural fit for each annotated sequence. All of the predicted BLUF sequences were aligned for an analysis of conservation and variation of residues using the Clustal Omega program.

3. Results and Discussion

3.1. BLUF Sequences, Modular Domains and Phylogenetic Analysis

We assessed the residues that are conserved and important for the substrate specificity in the respective orthologs. The details for each protein sequence and domain architecture are given in Table 1 and Figure 1.
Table 1. Blue light using flavin (BLUF) modular domains from different organisms.

<table>
<thead>
<tr>
<th>Domains</th>
<th>Accession No.</th>
<th>BLUF Length</th>
<th>Score</th>
<th>Organisms</th>
<th>Probable Modulations</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLUF + EAL</td>
<td>ARH96915.1</td>
<td>2–91</td>
<td>e&lt;sup&gt;−28&lt;/sup&gt;</td>
<td><em>Escherichia coli</em></td>
<td>Regulate diguanylate cyclases and Phosphodiesterase activity</td>
</tr>
<tr>
<td>BLUF + PsIE</td>
<td>EGW22399.1</td>
<td>2–92</td>
<td>e&lt;sup&gt;−36&lt;/sup&gt;</td>
<td><em>Methylobacter tundripaludum</em> SV96</td>
<td></td>
</tr>
<tr>
<td>BLUF + CHD</td>
<td>EHQ08139.1</td>
<td>3–93</td>
<td>e&lt;sup&gt;−28&lt;/sup&gt;</td>
<td><em>Leptotena illini</em> DSM 21528</td>
<td>Regulate adenyllyl and guanylyl cyclase activity</td>
</tr>
<tr>
<td>BLUF + B&lt;sub&gt;12&lt;/sub&gt; Binding domain</td>
<td>ABP71929.1</td>
<td>17–106</td>
<td>e&lt;sup&gt;−35&lt;/sup&gt;</td>
<td><em>Rhodobacter sphaeroides</em> ATCC 17025</td>
<td>Broadens BLUF photosensing ability</td>
</tr>
<tr>
<td>BLUF + PRK09039 superfamily</td>
<td>WP_012321331.1</td>
<td>5–99</td>
<td>e&lt;sup&gt;−36&lt;/sup&gt;</td>
<td><em>Methylbacterium radiotolens</em></td>
<td>Regulate phosphoribulokinase, uridine kinase and pantothenate kinase</td>
</tr>
<tr>
<td>BLUF + DNA&lt;sub&gt;pol3&lt;/sub&gt;_gamma3 superfamily</td>
<td>AIQ92835.1</td>
<td>5–99</td>
<td>e&lt;sup&gt;−36&lt;/sup&gt;</td>
<td><em>Poxyrobacter sp.</em> P11F6</td>
<td>Regulate oxidative degradation of steroids, fatty acid and xenobiotics</td>
</tr>
<tr>
<td>BLUF + AcrR</td>
<td>WP_058726129.1</td>
<td>3–94</td>
<td>e&lt;sup&gt;−28&lt;/sup&gt;</td>
<td><em>Curtobacterium lactum</em></td>
<td>Regulate antibiotic resistance in bacteria</td>
</tr>
<tr>
<td>BLUF + TetR&lt;sub&gt;C&lt;/sub&gt;_6</td>
<td>WP_051596720.1</td>
<td>4–92</td>
<td>e&lt;sup&gt;−34&lt;/sup&gt;</td>
<td><em>Curtobacterium sp.</em> UNCL17</td>
<td>Regulate antibiotic resistance in bacteria</td>
</tr>
<tr>
<td>BLUF + Endonuclease&lt;sub&gt;NS&lt;/sub&gt;</td>
<td>WP_058743091.1</td>
<td>17–83</td>
<td>e&lt;sup&gt;−30&lt;/sup&gt;</td>
<td><em>Drosophila rectora</em></td>
<td>Modulate hydrolase activity, nucleic acid and metal ion binding</td>
</tr>
<tr>
<td>BLUF + PAS</td>
<td>WP_058511962.1</td>
<td>4–88</td>
<td>e&lt;sup&gt;−13&lt;/sup&gt;</td>
<td><em>Legionella steels</em></td>
<td>Regulate cellular signaling processes</td>
</tr>
<tr>
<td>BLUF + AraC</td>
<td>WP_053973760.1</td>
<td>228–318</td>
<td>e&lt;sup&gt;−19&lt;/sup&gt;</td>
<td><em>Polaribacter dokdonensis</em></td>
<td>Control synthesis of structural components of arabinose metabolism</td>
</tr>
<tr>
<td>BLUF + Abhydrolase superfamily</td>
<td>XP_008692928.1</td>
<td>58–99</td>
<td>e&lt;sup&gt;−15&lt;/sup&gt;</td>
<td><em>Ursus maritimus</em></td>
<td>Modulate hydrolytic enzyme activity</td>
</tr>
<tr>
<td>BLUF + ENDO3c Superfamily</td>
<td>EYD78138.1</td>
<td>1–89</td>
<td>e&lt;sup&gt;−27&lt;/sup&gt;</td>
<td><em>Rubellimicrobium mesophilum</em> DSM 19309</td>
<td>Control DNA repair regulation</td>
</tr>
<tr>
<td>BLUF + ANK</td>
<td>EYJ08076.1</td>
<td>547–393</td>
<td>e&lt;sup&gt;−13&lt;/sup&gt;</td>
<td><em>Oxytricha triflax</em></td>
<td>Modulate protein-protein interaction</td>
</tr>
<tr>
<td>BLUF + DUF1115 Superfamily</td>
<td>WP_0229622806.1</td>
<td>5–101</td>
<td>e&lt;sup&gt;−37&lt;/sup&gt;</td>
<td><em>Pseudomonas pelagia</em></td>
<td></td>
</tr>
<tr>
<td>BLUF + RhoGEFSuperfamily</td>
<td>XP_025342216.1</td>
<td>171–262</td>
<td>e&lt;sup&gt;−38&lt;/sup&gt;</td>
<td><em>Pseudomicrostrea glucosiphilum</em></td>
<td>Control activation of Rho family GTPases</td>
</tr>
<tr>
<td>BLUF + PDZ</td>
<td>jgi_Bigna1_8555</td>
<td>55–144</td>
<td>e&lt;sup&gt;−17&lt;/sup&gt;</td>
<td><em>Bigelowiella natans</em> CCMP 2755</td>
<td>Regulate membrane-bound cell signaling</td>
</tr>
<tr>
<td>BLUF + AANH_like Superfamily</td>
<td>jgi_Schag1_101311</td>
<td>273–366</td>
<td>e&lt;sup&gt;−24&lt;/sup&gt;</td>
<td><em>Schizochytrium aggregatum</em> ATCC 28209</td>
<td></td>
</tr>
<tr>
<td>BLUF + EAL + GGDEF</td>
<td>AFL74487.1</td>
<td>456–546</td>
<td>e&lt;sup&gt;−33&lt;/sup&gt;</td>
<td><em>Thioctia violascens</em> DSM 198</td>
<td>Regulation of c-di-GMP level</td>
</tr>
<tr>
<td>BLUF + EAL + PRK15043 superfamily</td>
<td>CDW60191.1</td>
<td>2–62</td>
<td>e&lt;sup&gt;−25&lt;/sup&gt;</td>
<td><em>Trichurus trichurus</em></td>
<td></td>
</tr>
<tr>
<td>BLUF + GGDEF + PAS</td>
<td>WP_058465269.1</td>
<td>1–91</td>
<td>e&lt;sup&gt;−25&lt;/sup&gt;</td>
<td><em>Legionella cincinnatiensis</em></td>
<td></td>
</tr>
<tr>
<td>BLUF + GGDEF + COG5001 Superfamily + PAS</td>
<td>WP_058516015.1</td>
<td>3–85</td>
<td>e&lt;sup&gt;−23&lt;/sup&gt;</td>
<td><em>Legionella santarcucis</em></td>
<td></td>
</tr>
<tr>
<td>BLUF + COG5001 + PRB1_NHase + PAS</td>
<td>ADC61983.1</td>
<td>974–1064</td>
<td>e&lt;sup&gt;−38&lt;/sup&gt;</td>
<td><em>Alleochromatium vinosum</em> DSM 180</td>
<td></td>
</tr>
</tbody>
</table>
Table 1. Cont.

<table>
<thead>
<tr>
<th>BLUF + PSIE + RaeS Superfamily</th>
<th>WP_014418160.1</th>
<th>3–93</th>
<th>e&lt;sup&gt;-42&lt;/sup&gt;</th>
<th>Methyloribulokinase</th>
<th>NP*</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLUF + CHD + LRR_R1 Superfamily</td>
<td>Q8S9F2.1</td>
<td>57–136</td>
<td>e&lt;sup&gt;-24&lt;/sup&gt;</td>
<td>468–554</td>
<td>e&lt;sup&gt;-27&lt;/sup&gt;</td>
</tr>
<tr>
<td>BLUF + CHD + Med 26, M Superfamily</td>
<td>XP_03758351.1</td>
<td>98–186</td>
<td>e&lt;sup&gt;-19&lt;/sup&gt;</td>
<td>673–756</td>
<td>e&lt;sup&gt;-18&lt;/sup&gt;</td>
</tr>
<tr>
<td>BLUF + PRK11633 + DNA pol3 gamma3 family</td>
<td>WP_048452447.1</td>
<td>5–99</td>
<td>e&lt;sup&gt;-24&lt;/sup&gt;</td>
<td>Methylobacterium tarahaniae</td>
<td>NP*</td>
</tr>
<tr>
<td>BLUF + TefR_C, 11 family + Fer2_2 superfamily</td>
<td>WP_058743091.1</td>
<td>4–92</td>
<td>e&lt;sup&gt;-24&lt;/sup&gt;</td>
<td>Curtobacterium citreum</td>
<td>NP*</td>
</tr>
<tr>
<td>BLUF + GAL4 + Fungal TF MHR</td>
<td>ORY86082.1</td>
<td>10–100</td>
<td>e&lt;sup&gt;-38&lt;/sup&gt;</td>
<td>Protomycetes inouyei</td>
<td>Regulate galactose induced genes</td>
</tr>
<tr>
<td>BLUF + SRPBCC Superfamily + RGS Superfamily</td>
<td>BAV14116.1</td>
<td>518–612</td>
<td>e&lt;sup&gt;-26&lt;/sup&gt;</td>
<td>Naegleria fowleri</td>
<td>NP*</td>
</tr>
<tr>
<td>BLUF + SRPBCC + BTB + DUF3522 Superfamily</td>
<td>EFC49155.1</td>
<td>1132–1226</td>
<td>e&lt;sup&gt;-28&lt;/sup&gt;</td>
<td>Naegleria gruberi</td>
<td>NP*</td>
</tr>
<tr>
<td>BLUF + SRPBCC + FH2 + Drf_FH1 + PRK13729 + SMC_N</td>
<td>XP_002669619.1</td>
<td>1971–2050</td>
<td>e&lt;sup&gt;-28&lt;/sup&gt;</td>
<td>Naegleria gruberi</td>
<td>NP*</td>
</tr>
</tbody>
</table>

BLUF- Blue light using flavin; EAL- Glutamine/Alanine/Leucine; PRK- Phosphoribulokinase; PAS- Per/Ara/3Sim; PBP- Phosphatidylethanolamine-Binding Protein; CHD- Cyclase homology domain; LRR- Leucine-rich repeats; Med26- Mediator of RNA polymerase II transcription subunit 26; B2- Vitamin B2; DNA pol-DNA polymerase; REC- cheY-homologous receiver domain; HTH-Helix turn helix; Endo3- Endonuclease 3; ANK- Ankyrin repeats; RGS- Regulator of G protein signaling; DUF- Domain of unknown function; RhoGEF- Guanine nucleotide exchange factor for Rho/Rac/Cdc42-like GTPases; GAL4- GAL4-like Zn(II)2 Cys6 (or C6 zinc) binuclear cluster DNA-binding domain; MHR- Middle homology region; PDZ- PSD95/Dlg1/zo-1; SRPBCC-START/RHO_alpha_C/PITP/Bet_v1/CoxG/CalCligand-binding; BTB- Broad-Complex, Tramtrack and Bric a brac; FH2- Formin Homology 2; Drf- Diaphanous related forms; SMC_N- N terminus of structural maintenance of chromosomes. Sequences WP_05159620.1 from Curtobacterium sp. UNCC117; Q8S9F2.1 from Euglena gracilis; and XP_013758351.1 from Thecamonas trahens ATCC 50062 are strongly predicted to contain two BLUF domains on either side of their respective effector domains. NP*: Probable modulated function cannot be predicted, but requires a detailed study.  

Figure 1. Schematic representation of the different blue light using flavin (BLUF) modular domain containing proteins. The accession numbers were taken from National Center for Biotechnology Information (NCBI). The mentioned “AA” indicates the amino acid numbers of the particular BLUF modular protein.
The analysis of 34 BLUF sequences revealed that residues forming BLUF catalytic core are conserved throughout for interaction with the flavin chromophore (Figure 2). The amino acids tyrosine (Y), asparagine (N), glutamine (Q), and tryptophan (W) or methionine (M), which are crucial for the photodynamics and photocycle of the BLUF domain, are highly conserved (Figure 2). Our analysis further confirms that tyrosine, glutamine, and tryptophan (or methionine) are critical for the substrate specificity of the BLUF sequences, as has been reported earlier [6,11,33,57]. The photo-activation of the BLUF domains actually involves conserved glutamine and tyrosine residues, where glutamine contributes to hydrogen bond formation in a dark state [11–14]. However, in light adapted conditions, a hydrogen bond rearrangement (tautomerization) occurs to form a new hydrogen bond with tyrosine [11]. A motif analysis revealed three different conserved motifs (Figure 2) among different BLUF sequences. Each motif has its importance, as each of them was comprised of an essential amino acid residue involved in the regulation of the BLUF photocycle and photodynamics [6,11].

**Figure 2.** Multiple sequence alignment of the different BLUF modular domains depicting conserved amino acids. The black arrow indicates conserved amino acids crucial for regulating the flavin binding pocket, photocycle and photodynamics of the BLUF domain containing proteins [6,11]. The sequences under the solid boxes represent the conserved motifs of the BLUF domain. The conserved motifs were predicted using the Multiple EM for the Motif Elicitation (MEME) suite.

The phylogenetic analysis of the selected sequences of 34 BLUF domains divides them into four distinct clusters (I–IV), which seem to differ from each other mainly by the presence of eukaryotic and prokaryotic BLUF proteins in them (Figure 3). As observed from the phylogenetic analysis, the BLUF domain sequences of eukaryotic and prokaryotic origins are evolutionarily intermixed. Each cluster (except cluster IV) was further subdivided into sub-clusters, which represent closely related BLUF domain sequences broadly associated with similar kind of effector domains (Figure 2).
Figure 3. Phylogenetic analysis by the Maximum Likelihood (ML) method. The analysis was done using 34 amino acid sequences of the modular BLUF domain containing proteins. All positions containing gaps and missing data were eliminated. There were a total of 17 positions in the final dataset. A solid red circle represents protein sequences of prokaryotic origin, and a solid red square represents protein sequences of eukaryotic origin. Evolutionary analyses were conducted using MEGA6 [51].

3.2. Modular Diversity of BLUF Domains

In the present communication, we have selected 34 different modular architectures for the BLUF domain in combinations with various effector domains (Table 1; Figure 1). The annotated BLUF domains in our database sequences range from 80 to 100 amino acids, situated mostly at the N-terminal region of the effector domains. However, the predicted BLUF sequences in our database show a high sequence similarity across the BLUF region, with E-values in the range of e\textsuperscript{−40} to e\textsuperscript{−120}. The residues show an upper conservation pattern around the flavin binding region (Figure 4). The secondary structural analysis indicates that the BLUF sequences are composed mostly of five \(\beta\) strands and two \(\alpha\) helices. The E-value, sequence length and sequence conservation details are summarized in Table 1. Three-dimensional models of the BLUF domains were constructed using the Phyre2 modeling software [53]. The BLUF domains reveal conserved \(\beta\alpha\beta\alpha\beta\beta\) fold conformations in which two \(\alpha\)-helices surround a five-stranded antiparallel \(\beta\)-sheet platform (Figure 4).
Figure 4. Models of the BLUF domains and details of its flavin binding pocket. (A) Superposition of the modelled BLUF domains using the Phyre server with a crystal structure of *Rhodobacter sphaeroides* BlrB (PDB: 2BYC) (magenta). The homology models include the annotated BLUF domains from WP_045444510.1 (*Psychrobacter sp.*) (Green tone); WP_014148160.1 (*Methylomicrobium alcaliphilum*) (Brown tone); XP_025342216.1 (*Pseudomicrostroma glucosiphilum*) (Blue tone); AFL74487.1 (*Thiocystis violascens*) (Violet tone); EHQ08139.1 (*Leptonema illini*) (Yellow tone); ABP71929.1 *Rhodobacter sphaeroides* ATCC 17025 (Pink tone); WP_0229622806.1 (*Pseudomonas pelagia*) (Skyblue tone); and ORY86082.1 (*Protomyces inouyei*) (Smudge tone). The domain boundaries of the modelled BLUFs are mentioned in Table 1. (B) The BLUF photocycle scheme shows the protein/FAD interactions through the hydrogen bonding pattern of the flavin moiety with conserved glutamine upon illumination. (C) Superposition of the flavin binding pocket in the BLUF models in comparison to *Rhodobacter sphaeroides* BlrB (PDB: 2BYC) and the BLUF domain of AppA (PDB: 1YRX). The side chains of residues with potentially important roles in catalysis and/or substrate binding are shown as stick models and are labelled. The selected regions of the same are shown in a reduced multiple sequence alignment.

We modeled the binding modes of the flavin chromophore based on the observations in the known BLUF domain structures [33,57]. The isoalloxazine ring of flavin can be readily accommodated in the pocket in the models sandwiched between two α helices, suggesting that the active site is appropriately formed in BLUF models (Figures 4 and 5). Spectroscopic analyses and photochemistry
of BLUF proteins have shown that flavin chromophore facilitates blue light-induced electron transfers by a hydrogen bond rearrangement between flavin N5 and O4 and conserved tyrosine, glutamine, and tryptophan or methionine of the BLUF [14,17], leading to a spectral shift from blue to red and the transmission of the signal further downstream. Sequence alignment studies of 34 BLUF domain containing sequences, in this study and elsewhere, show the conservation of tyrosine and glutamine residues in the β1 and β3 strands of the BLUF fold (Figure 4), and this has been reported to be critical for the photochemical reaction, as mutations in these amino acids result in the loss of its ability to perceive light [12–14,17,58]. In our sequence database, BLUF photo domains are fused at the N-terminus of a wide array of different effector domain modules, viz. kinases, phosphatases, phosphodiesterases, anti-sigma factors, DNA binding domains, a transcriptional repressor (TetR), PAS, endonuclease and PBP proteins. Some of these have been modeled with >90% confidence using the Phyre2 modeling program and are presented in Figure 4. These domains are part of various photoreceptors [1,59,60]. Our structural and field mapping analysis also shows that sequences like TetR_C_6, SRPBC1 superfamily, GGDEF1 COG5001 superfamily, COG5001 super family PBP1_NHase photoreceptors [1,59,60]. Our structural and field mapping analysis also shows that sequences like TetR_C_6, SRPBC1 superfamily, GGDEF1 COG5001 superfamily, COG5001 super family PBP1_NHase photoreceptors [1,59,60]. Our structural and field mapping analysis also shows that sequences like TetR_C_6, SRPBC1 superfamily, GGDEF1 COG5001 superfamily, COG5001 super family PBP1_NHase photoreceptors [1,59,60]. Our structural and field mapping analysis also shows that sequences like TetR_C_6, SRPBC1 superfamily, GGDEF1 COG5001 superfamily, COG5001 super family PBP1_NHase photoreceptors [1,59,60]. Our structural and field mapping analysis also shows that sequences like TetR_C_6, SRPBC1 superfamily, GGDEF1 COG5001 superfamily, COG5001 super family PBP1_NHase photoreceptors [1,59,60].

Figure 5. Representation of 3D structural models of BLUF (Red) and effector domain combinations (Blue) using the Phyre server [54]. (A) WP_0229622806.1 (Pseudomonas pelagia) (206 aa) codes a combination of the BLUF (5–101) and DUF domain (99–181). The model was generated with 100% confidence covering 1–186 residues using the R. sphaeroides AppA (PDB: 4HH0) as the template. (B) WP_051596720.1 (Curtobacterium sp. UNCCL17) (470 aa) codes for BLUF in combination with the transcriptional repressor (224–327). The model was generated with >90% confidence covering 1–350 residues using 12 different templates. (C) Q8S9F2.1 (Euglena gracilis) (1019 aa) codes for BLUF in combination with cyclase homology domains (CHDs), which are part of the class III nucleotidylyltransferases (20–379). The model was generated with >90% confidence covering 1–800 residues using 13 different templates. (D) ARf96915.1 (Escherichia coli) (403 aa) codes for BLUF (2–93) in combination with the EAL signaling domain (150–389). The model was generated with 100% confidence covering 1–389 residues using K. pneumoniae BrP1 (PDB: 3GFZ) as the template. The structures are represented as interactive coloured ribbons. The model images were generated using PyMol (http://www.pymol.org) [63].
3.3. BLUF Modules in Association with the Effector Domains

The occurrence of the BLUF domain in different associations reveals the diversity and abundance of this modular domain in a wide range of organisms. Our sequence and fold analyses confirm various types of effector modules fused with the BLUF domains (Figure 5).

3.3.1. EAL and GGDEF Domain

The EAL and GGDEF domain-containing proteins are widely distributed among bacteria and are involved in the regulation of the cellular level of the universal signaling molecule bis-(3',5')-cyclic-guanosine monophosphate (c-di-GMP), where the former act as diguanylate cyclases (DGCs) while the latter ones are phosphodiesterases (PDEs) [64]. The c-di-GMP generally controls a variety of signaling pathways associated with cell differentiation, bacterial adhesion and biofilm formation, bacterial motility, the colonization of host tissues and virulence [65,66]. In the EAL-GGDEF domain containing proteins, the GGDEF and EAL motifs in the active sites are crucial for the DGC and PDE enzyme activities [67–69]. Although most of the proteins involved in c-di-GMP signaling contain the GGDEF/EAL domains as a single polypeptide, possessing both the DGC and PDE enzyme activities, in some cases the GGDEF and EAL domain is also found alone [68]. In some of the EAL-GGDEF domains containing proteins, it was also reported that one [70,71] or both [72,73] of these domains are catalytically inactive due to a lack of respective GGDEF and EAL motifs that are crucial for the enzymatic function. In these proteins, the inactive domains either act as regulators [74] or c-di-GMP effectors [72,73]. Yang and coworker have reported a FimX-like protein (Flip) with a degenerate EAL-GGDEF domain which interacts with the PilZ-Domain protein to control virulence in Xanthomonas oryzae pv. oryzae [75]. YhdA is another protein with a degenerate EAL-GGDEF domain that promotes the turnover of CsrB and CsrC (small RNAs), which reduce the expression of the flhDC (flagellar master regulator) by sequestering the CsrA (RNA-binding protein) [76]. Several EAL-GGDEF domains containing proteins are also reported to have N-terminal sensory domains that can regulate the GGDEF and/or EAL domain functions. In the present study, we have selected an EAL only and EAL-GGDEF domain-containing proteins with the blue light using flavin (BLUF) domain as a sensory domain from E. coli and Thiocystis violascence, respectively (Table 1).

Barends and coworker have characterized the Klebsiella pneumoniae BlrP1 photoreceptor protein biochemically, structurally, mechanistically, and have elucidated the mechanism of the light-induced regulation of the EAL domain (PDE) via the BLUF sensor domain [4]. Upon light illumination, the structural change in the flavin binding pocket (Trp replace Met) of the BLUF domain cross-activates the EAL domain via allosteric communication and increases the PDE activity [4]. The photo-dependent alteration of the BLUF–EAL interactions influences the quaternary structure, the EAL–EAL interface at the dimerization helix, the compound helix \( \alpha_5^{EAL} \) and the loop connecting it to \( \beta_5^{EAL} \), which are implicated in the EAL activation [4,77]. In the BlrP1 photoreceptor, the BLUF domain shares a similar architecture, as shown by other BLUF domains from different organisms, where the central BLUF domain (N-terminus) is surrounded by two helices (helical cap; C-terminus) [78–80]. However, an additional EAL output domain has been identified in the BlrP1 protein, which is connected to the BLUF domain via a 50 Å long linker peptide (triosephosphate isomerase (TIM)-barrel fold) [4]. The EAL active site in the BlrP1 photoreceptor involved Glu188, Asn239, Glu272, Asp302, Asp303, Lys323 and Glu359 [4]. We used this well-characterized BlrP1 as a template and aligned it pairwise with EAL only (E. coli), EAL-GGDEF (Thiocystis violascence DSM 198), and EAL-PRK15043 domains (Trichuris trichuris), and observed that most of the amino acids constituting the EAL active site in BlrP1 are highly conserved through all of the EAL domain-containing proteins (Figure S1), which suggested that the EAL domain-containing proteins possess a similar mechanism for the PDE activity.

The query protein sequence (ARH96915.1) containing the EAL output domain was subjected to a protein-protein interaction analysis, which revealed several interacting partners involved in the regulation of different signaling pathways (Figure S7a; Table S1). The protein-protein interaction analysis showed that most of the interacting partners for this photoreceptor are either EAL domain-containing
PDEs (JD73_03740) or GGDEF domain-containing DGCs (JD73_23675, JD73_25605, JD73_23680, YeaP, and YdaM). However, the interactions with YcgZ (two-component connector protein) and YcgE (Mer-like repressor protein/transcriptional regulator) have also been revealed, indicating their involvement in the regulation of bacterial biofilm formation [30]. Tschowri et al. [30] have characterized the previously unknown function of the BLUF-EAL domain-containing protein, YcgE, from *E. coli*, and suggested that upon blue light irradiation, this protein acts like an antirepressor. The antirepressor YcgF removes YcgE (Mer-like repressor) from the promoter, and resumes the expression of different small regulatory proteins (YmgA and YmgB). These small regulatory proteins (YmgA and YmgB) utilize the RcsC/RcsD/RcsB two-component phosphorelay system to activate the production of colanic acid, a biofilm matrix component, and to decrease adhesive curli fimbriae [30]. The query protein may also interact with AriR/YmgB (regulator of acid resistance influenced by indole), another biofilm-related protein involved in the regulation of acid resistance in *E. coli* [81]. A string analysis also revealed the possible interaction between the query protein and regulatory protein, LuxR, which regulates quorum sensing in the bacterial system [82,83].

### 3.3.2. PsiE Domain

In *E. coli*, the phosphate-starvation-inducible (*psiE*) gene is positively and negatively regulated by both PhoB and cAMP-CRP (cAMP receptor protein), which are respectively involved in the phosphate and carbon metabolism [84]. The phosphate and carbon sources regulate the *psiE* gene by using the *lacZ* and *chloramphenicol acetyltransferase* gene (*cat*) fusions, respectively [84]. Although the function of PsiE has not yet been determined, sometimes it has been predicted to have features like DNA-binding protein inhibitor-related, putative transcriptional regulators or hypothetical DNA binding proteins (IPR020948).

#### 3.3.3. Cyclase Homology Domain (CHD)

The cyclase homology domains (CHDs) are the catalytic domains of eukaryotic and prokaryotic nucleotidylcyclases, i.e., adenylyl cyclases (ACs) and guanylyl cyclases (GCs), which belong to the evolutionary diverse class III nucleotidylcyclases. CHDs are reported as three different structural forms, i.e., heterodimers (mammalian CHDs), pseudoheterodimers (metazoan CHDs) and homodimers (bacterial and protozoan CHDs) [85]. Heterodimeric and pseudoheterodimeric CHDs have a single catalytic pocket sharing catalytic amino acid residues at the dimer interface, while homodimeric CHDs have two separate catalytic pockets, each of which contributes the CHD determinant [85]. In spite of having two potential catalytic pockets, several enzymes with homodimeric CHDs (for example, eukaryotic class III nucleotidylcyclases) may have only one catalytically competent site [86]. Although CHDs are structurally diverse, all of them have a conserved structural component (i.e., a helical region) mutating which compromises the stability and active dimeric conformation of a protein [87]. All CHDs have a common catalytic mechanism in which they require two magnesium or manganese ions to bind a polyphosphate group of the nucleotide, followed by nucleophile activation.

Most CHDs, except a few (Rv1359 from *M. tuberculosis*), are reported to exist in combination with different regulatory modules, thus making it possible to perceive the variety of signals and the regulation of the intracellular cAMP generation [85]. In the present communication, we have selected three such proteins, with the BLUF domain in combination with the CHD domain, having the accession numbers EHQ08139.1 (from *Leptomonas illini* DSM 21528), Q8S9F2.1 (from *Euglena gracilis*) and XP_013758351.1 (from *Thecamonas trahens* ATCC 50062) (Table 1). In *Euglena gracilis*, CHDs (adenylyl cyclases) occur in combination with the BLUF domain, with an overall domain arrangement of BLUF1CHD1BLUF2CHD2, where the flavin chromophore senses blue light and stimulates the adenylyl cyclase activity [35]. Furthermore, we have performed a alignment of sequences representing the CHD domain from selected proteins against the progression of the well-characterized template CHD domain of bPAC (Figure S2). The multiple sequence alignment revealed that active site residues involved in the nucleotide binding (i.e., Asn257-His266, Lys263-Met264 (forming the β4AC-β5AC tongue), Gly259-Asn178 (forming the α2AC helix) and Lys263–Thr196, Asp265–Phe198, and His266–Lys197 (forming the β2AC-β3AC
hairpin) [33]) are highly conserved among all three selected BLUF-CHD domain-containing proteins (Figure S2). In PACs, under dark conditions, the orientation and arrangement of an individual amino acid in the active site (i.e., Thr267 and Lys197 (bound to Phe198)) render the conformation of AC inactive. On the other hand, upon illumination, the change in the interaction between Asn25 and His266 resulted in the correct orientation of Thr267 required for the communication with the adenine base; additionally, the Lys197 is detached from Phe198, thus providing space for the adenine base to enter the active site more deeply [33]. We also performed a protein-protein interaction analysis, which revealed the possible interacting partners for CHD domain-containing proteins, which range from phosphodiesterases, the RNA polymerase subunit β, and the DNA helicase to another adenylate cyclase/guanolate cyclase associated with the GAF and PAS/PAC sensor (Figure S7b and Table S1).

3.3.4. PAS Domain

PAS (Per-Arnt-Sim) domain-containing proteins are widely distributed among all domains of life. PAS domain acts as a sensor, generally found at the N terminus of sensory and signaling transduction related proteins, and detect a variety of stimuli and regulating the functions of a diverse array of effector domains [88,89]. Members of the PAS domain family can bind a diverse range of small-molecule metabolites [22], which could either directly act as a signal and be involved in initiating a cellular signaling response [90], or which could serve as a cofactor and respond to subsequent messages like gas molecules, redox potential, or photons [39]. Although PAS domain-containing proteins are chemically and functionally diverse, almost all PAS domains have a conserved core comprised of a five-stranded antiparallel β-sheet and several α-helices, which are responsible for the generation and propagation of a signal to the adjoining effector domain. In the present study, we have selected four proteins, three of them from different Legionella strains and one from Allochromatium vinosum DSM 180, respectively, having combinations of the BLUF and PAS sensor domain (Table 1). We performed the alignment of sequences representing the PAS domain in four different proteins against a well-characterized photoactivated yellow protein (PYP; 1NW_Z) from Halorhodospira halophila [91] (Figure S3). As discussed earlier, PAS domains are structurally diverse; the same is revealed from the multiple sequence alignment analysis. Although, the PAS core (5’ NAAEGDIT 3’) in PYP is not conserved among other PAS domain-containing proteins, interestingly, in three of the selected PAS domains containing proteins for Legionella genus, the sequences representing the PAS core are conserved (Figure S3). From the above observation, we could suggest that, although they are diverse in different organisms, PAS cores might be conserved in plants belonging to the same genus. We also performed a protein-protein interaction analysis, which revealed the possible interacting partners for PAS domain-containing proteins, which range from multisensor histidine kinase, CheA signal transduction histidine kinase, CheW protein, CheB methylesterase, 4-coumarate-CoA ligase, phenylalanine/histidine ammonia-lyase and Hpt sensor hybrid histidine kinase (Figure S7c and Table S1).

3.3.5. B₁₂ Binding Domain

Many prokaryotes synthesize vitamin B₁₂ (cobalamin) having a tetrapyrrole-like structure composed of a bound cobalt atom (Co) with two axial ligands. The lower ligand is known to be involved in vitamin B₁₂ binding, while the upper one is used as a cofactor for different groups of enzymes/proteins, such as methyltransferases, reductases and isomerases [92]. The vitamin B₁₂ binds to a specific domain, i.e., Asp/Glu-X-His-X-X-Gly-(41)-Ser/Thr-X-Leu-(26–28)-Gly-Gly, which is highly conserved in almost all B₁₂-dependent enzymes/proteins and characterized as a Rossmann fold, typically composed of 5 parallel β-sheets surrounded by 4–5 α helices [93,94]. Generally, vitamin B₁₂ and its derivatives are known for their role in fatty acid and folate metabolism; however, recently they have been characterized as photoreceptors with a novel and unanticipated biological function as a light dependent transcriptional regulator [95]. Ortiz-Guerrero et al. [96] reported, for the first time, a light-induced excitation of CarH (Mer-like transcriptional factor/repressor) bound adenosylcobalamin (AdoB₁₂), which inhibited the formation of a stable CarH-AdenoB₁₂ tetramer, thus allowing the gene expression
for the carotenoid biosynthesis in *Myxobacteria*. In the dark, the stable CarH-AdoB$_{12}$ tetramer binds at the promoter region and shuts down the carotenoid biosynthesis [95–97]. Upon light illumination of the corrin ring associated with AdoB$_{12}$, it promotes the re-orientation of a helix bundle forming a covalent linkage between H132 and Co, and causes the CarH dissociation from the promoter region, which ultimately leads to the carotenoid gene expression [95,96,98]. Cheng et al. [92] also reported a small stand-alone B$_{12}$-binding domain protein, AerR in *Rhodobacter capsulatus*, which controls the light-dependent regulation of the biosynthesis of the photosystem via interacting with CrtJ, a repressor of the photosystem gene expression. Like CarH, the light illumination also leads to a covalent association between the His10 and Co ligand, which suggested that the light-dependent covalent linkage between the Co ligand and His residue might be the common mechanism in B$_{12}$-dependent photoreceptors. Furthermore, photoreceptors have also been naturally endowed with multiple photosensory domains, which could relay signals to output domains to control specific light-dependent functions [95]. In the present communication, we have selected a modular protein (accession number ABP71929.1) with the BLUF and vitamin B12 binding domain from *Rhodobacter sphaeroides* ATCC 17025 (Table 1).

We performed the alignment of the selected protein with the well-characterized B$_{12}$ binding domain containing proteins (CarH and AerR) and, surprisingly, observed that the crucial amino acids (Trp131, Val138, Glu141 and His142) involved in forming the binding pocket for AdoB$_{12}$ [95,99] were not observed in the selected protein sequence (Figure S4). Moreover, there are many prokaryotes which do not necessarily require the B$_{12}$ cofactor; hence, they can acquire alternative B$_{12}$-independent metabolic pathways for the same reaction [100]. The association of the B$_{12}$ domains with the BLUF domains was also reported by Cheng et al. [92], where the role of the B$_{12}$ domain is to sense light that is out of the absorption range of flavin. In several proteins, B12 binding domains are also found in combination with heme/oxygen sensing globin domains; however, none of these proteins have been characterized [92]. In proteins with combinations of histidine kinases or serine/threonine kinases with the B$_{12}$ domain, the kinases are responsible for regulating the response to light absorption by the B$_{12}$ domain [93].

The selected protein sequence (ABP71929.1) containing the B$_{12}$ domain was analyzed for a protein-protein interaction using String (version 11). The observation showed several interacting partners involved in the regulation of different signaling pathways (Figure S7d). The protein-protein interaction analysis revealed the interaction of the B$_{12}$ domain with the prephenate dehydratase enzyme (Rsph17025_0524; Figure S7d), which catalyzes the conversion of prephenate to phosphoenolpyruvate (PEP), water, carbon dioxide [101], and is generally involved in phenylalanine, tyrosine and tryptophan biosynthesis [102,103]. The selected protein also interacted with tyrosyl-tRNA synthetase (TyrS), an enzyme that catalyzes the attachment of tyrosine to tRNA in a two-step reaction. The interaction of the B$_{12}$ domain-containing protein with a transmembrane protein PA-phosphatase-like phosphodiesterase (Rsph17025_2732), a SARP family transcriptional regulator and a DNA mismatch repair protein, MutL, was also revealed through the protein-protein interaction analysis (Figure S7d). The BLUF-B$_{12}$ binding domain containing protein also interacted with the TonB protein, which communicates with outer membrane receptor proteins and drives the energy-dependent uptake of various substrates (such as iron citrate, enterochelin, aerobactin, etc.) into the periplasmic space [104].

3.3.6. PRK Superfamily

The PRK family represents a group of three types of P-loop containing kinases, i.e., phosphoribulokinase [105], uridine kinases [106], and pantothenate kinases (CoaA) [107]. This family is named after one of its members, i.e., phosphoribulokinase (PRK), which drives the phosphoryl transfer from Mg-ATP to ribulose 5-phosphate to form ribulose 1, 5-bisphosphate (RuBP) during the pentose phosphate pathway [108]. In *E. coli*, the *udk* gene encodes the pyrimidine salvage enzyme uridine kinase, causing the phosphorylation of uridine/cytidine into UMP/CMP using GTP as the phosphate donor [106]. Pantothenate kinase controls the rate-limiting step in the coenzyme A (CoA) biosynthesis pathway [107]. The *coaA* gene is transcribed to produce the 1.1 kb transcript, which is further translated.
into two protein products of 36.4 and 35.4 kDa, respectively [107]. The resultant proteins encoded by the coaA gene, showed a difference of eight amino acids at the N-terminus. The E. coli strains bearing multiple copies of the coaA gene showed a higher activity of the pantothenate kinase [107].

3.3.7. DNA pol 3 gamma3 Superfamily

In E. coli, DNA polymerase III (Pol III) is a complex holoenzyme comprised of three functionally distinct subassemblies, i.e., the core polymerase (α, ε, and θ subunit), the sliding clamp (β subunit) and the clamp loader complex (τ2γδ′χψ subunit) [108]. The clamp loader is responsible for the DNA-dependent hydrolysis of ATP to load β2 clamps onto DNA for the interaction with core polymerases [109]. The gene dnaX encodes the ATP motor subunits of the clamp loader, i.e., one γ and two τ subunits, where the γ subunits are considered as a truncated product of the τ subunits [110,111].

The gamma (γ) subunit shares domains I-III with the tau (τ) subunit, while the domain IV and the entire alpha-interacting domain V subunit are only observed in the τ-subunit. The bacterial DNA pol III γ III domain and its homolog, the eukaryotic replication factor C (RFC), belong to the AAA-ATPase superfamily and are primarily involved in the breaking or restructuring of the supramolecular assembly of proteins [110]. In this communication, we have selected two BLUF modular domains in association with the DNA pol III γ III domain only (accession number AIQ92835.1) from Methylobacterium oryzae CBMB20, and the PRK11633-DNA pol III γ III domain (WP_048452447.1) from Methylobacterium tarhaniae (Table 1).

Furthermore, we have performed the alignment of the sequences representing the DNA pol III γ III domain against the well-characterized truncated sequence (1-373 amino acids) of the DNA polymerase III subunit gamma/tau (WP_113440333.1) from E. coli (Figure S5). The sequences corresponding to the DNA pol III γ III domains aligned against the domain II of the γ III subunit of DNA pol III from E. coli; however, most of the critical amino acids are not conserved amongst the DNA pol III γ III domain. In the E. coli DNA pol III γ III domain, Thr157, responsible for the hydrogen bond formation with the terminal phosphate of AMP-PNP, is only conserved in the DNA pol III γ III domain from Methylobacterium tarhaniae but not in Methylobacterium oryzae CBMB20 (Figure S5) [110]. Moreover, the C-terminal SARC motif (Ser168, Arg169, and Cys170) located in the α7 helix in the sensor1 region, which is highly conserved in the γ subunit of almost all organisms [112,113], is found missing in both of the DNA pol III γ III domains. Arg169 has dual roles, where on the one hand it acts as an “arginine finger” for the SARC motif, while on the other hand it is responsible for electrostatic and hydrophobic interactions which hold the δ subunit onto the γ III subunit. Arg215 is another critical amino acid and is a part of the conserved motif G/PxΦRXΦ (where Φ is any hydrophobic residue) located in the DNA pol III γ III domains, among prokaryotes as well as among eukaryotes [112]. The correct alignment of Arg215 is very crucial for its proper interaction with the phosphate group of ADP/ATP [110]. From the multiple sequence alignment analysis, it was observed that this particular amino acid is conserved in the BLUF-DNA pol III γ III domains but replaced by another polar amino acid (tyrosine:Q) in the BLUF-PRK11633-DNA pol III γ III domain (Figure S5).

The protein-protein interaction analysis revealed that most of the interacting partners of DNA pol III γ III (dnaX) are the components involved in the regulation of DNA replication, such as DNA pol I (Pol A), replicative DNA helicase (dnaB), DNA mismatch repair protein (recR), DNA pol III subunit α (dnaE), δ (holA and holB), ε (dnaQ), chi (holC), psi (holD) and β sliding clamp (dnaN) (Figure S7e and Table S1).

3.3.8. Cytochrome p450/ p450 Superfamily

Cytochrome p450s (CYPs) are a diverse group of heme-containing monoxygenases responsible for the oxidative degradation of steroids, fatty acid and xenobiotics [114]. These heme-thiolate proteins are named after their spectral absorbance peak at 450 nm, due to linkage with the cysteine thiolate of the protein [114]. In spite of a low sequence conservation, the structures are highly conserved. The cytochrome p450 core is made up of a four-helix bundle, helices J and K, two sets
of beta-sheets, and it possesses a heam-binding loop, a proton-transfer groove and the conserved EXXR motif in helix K [115]. The hormone synthesis, cholesterol, and vitamin metabolism are some other pathways that are regulated by cytochrome p450s. In this study, we have selected a protein containing the BLUF domain associated with the p450 domain (accession number WP_045444510.1) from Psychrobacter sp. P11F6 (Table 1). A multiple sequence alignment of the sequence representing the p450 domain against the well-characterized CYP for Bacillus subtilis has been performed (Figure S6). The multiple sequence alignment revealed that, among two amino acids (i.e., arginine (Arg242) and proline (Pro243)), essential for substrate binding in the peroxygenase enzyme [116], arginine (Arg108) is also conserved in the BLUF regulated p450 domain (Figure S6). However, in the BLUF controlled p450 domain, hydrophobic proline is replaced by a polar amino acid residue, i.e., serine (Ser109). In several p450 enzymes, an adjacent acidic-polar amino acid pair was reported in the substrate binding site. In Pseudomonas putida camphor hydroxylase (CYP101A1), Asp251 and Thr252 were observed in the substrate binding site and used to relay protons onto iron-oxo species to activate the catalytic cycle [114]. The above observations indicate the different evolutionary routes adapted by these enzymes for the H2O2-driven catalysis. We also performed a protein-protein interaction analysis, which showed interacting partners for the selected protein (Figure S7f and Table S1). Most of the interacting proteins belonged to the fatty acid metabolism (CypB, CypC, CypD, and YitS) and to the secondary metabolism (PksJ and PksM).

3.3.9. REC Domain

Signal receiver (REC)/CheY-like photo-acceptor domains are the widely distributed regulatory domains in bacteria (CheY, OmpR, NtrC, and PhoB), however, they are now also reported in eukaryotes, for example, ETR1 from Arabidopsis thaliana. In the bacterial two-component regulatory system, the response regulator typically consists of a receiver domain that is covalently linked to an effector domain (DNA binding or catalytic units), and which is controlled by sensor kinase-catalyzed aspartyl phosphorylation [117]. The role of the REC domain is to receive the input signal perceived and transmitted from the sensor partner in the two-component systems. The REC domain interacts with different proteins to regulate processes like bacterial chemotaxis and some other regulatory pathways [118].

3.3.10. TetR and AcrR Domain

The TetR protein family is a group of transcriptional regulators with an HTH DNA-binding motif, which is widely distributed among bacteria [119]. TetR family proteins control efflux pumps and transporters having a role in antibiotic resistance and tolerance to toxic chemicals, synthesis of osmoprotectants, quorum sensing, drug resistance, virulence and sporulation [120,121]. However, the TetR family is named after its most characterized member, TetR, which has a role in the regulation of the expression of tet genes, involved in conferring tetracycline resistance in the bacterial system [120]. Proteins with the TetR domain maintain its optimal cellular level by feedback control. TetR (dimer) binds to two adjacent DNA major grooves (6bp each) located in the promoter region of the target gene on both of the strands, where helix α3 (Gln38 to His44) is involved in a sequence-specific recognition [119,120]. The Arg28 in helix α2 strengthens the specific contact with the complementary strand [119,120]. The hydrophobic core, developed from the contributing residues from the α1, α2, and α3 bundle, stabilizes the TetR DNA binding domain [119]. A highly conserved Lys48, located in α4, also has an essential role in the TetR-DNA complex formation [119].

The TetR family protein also works in a complex circuit with other proteins, including AcrR, a transcriptional repressor of the acrAB operon responsible for encoding a multidrug efflux pump which removes a wide range of antibiotics and confers antibiotic resistance in E. coli [121]. The AcrR protein (215 amino acid; dimer) crystal structure showed that it is composed of a three-helix DNA-binding domain and a unique C-terminal domain (large internal cavity) for ligand binding, which is structurally similar to members of the TetR family of transcriptional repressors [120,122]. It was predicted that
Appl. Sci. 2019, 9, 3924 16 of 28

ligand (rhodamine 6G, ethidium and proflavin) binding at the C-terminal ligand binding site leads to an alteration in the conformation of the N-terminal DNA binding region and thereby initiates transcription at the corresponding promoter of the target gene [123].

3.3.11. Endonuclease 3c and Endonuclease-NS Domain

The endonuclease 3c domain is widely distributed among the family of DNA repair proteins such as endonuclease III and DNA glycosylase (MutY or MBD4). The members of this family possess a conserved helix-hairpin-helix (HhH), a Gly/Pro-rich loop, as well as a conserved aspartate residue [124,125]. On the other hand, the endonuclease-NS domain has been explicitly reported in the DNA/RNA non-specific endonucleases and found both in prokaryotes and eukaryotes. The endonuclease-NS domain, containing endonucleases, showed an Mg\(^{2+}\) dependent cleavage of double-stranded as well as single-stranded nucleic acids. The extracellular *Serratia marcescens* nuclease is a well-characterized example of an endonuclease with an endonuclease-NS domain having a conserved histidine residue. The *Serratia marcescens* nuclease requires magnesium ion, three acidic (Asp107, Glu148 and Glu232) amino acid residues, as well as a few basic amino acid residues (Arg108, Arg152) for the endonuclease activity [126,127]. Proteins with the endonuclease-NS domain are broadly involved in hydrolase activity, nucleic acid binding and metal ion binding [126].

3.3.12. AraC Domain

The AraC protein (inducer/activator) regulates the *araBAD* operon in *E. coli*, which is responsible for encoding structural components for the arabinose metabolism [128]. X-ray crystallization and NMR studies demonstrated that the AraC is a dimeric protein composed of two helix-turn-helix DNA-binding motifs [129]. AraC uses arabinose as a substrate, and the induction of the *araBAD* operon depends on the concentration of extracellular arabinose (>10\(^{-7}\) M) as well as on the rate of the arabinose uptake and catabolism [128].

3.3.13. Abhydrolase (α/β Hydrolase) Superfamily

The α/β hydrolase superfamily is a diverse group of hydrolytic enzymes that may differ in their catalytic function but that share a common fold with a conserved loop bearing catalytic triad [130]. The catalytic triad involves serine, glutamate/aspartate, and a histidine amino acid residue, and participates in the nucleophilic attack on a carbonyl carbon atom. The α/β hydrolase fold includes proteases, lipases, peroxidases, esterases, epoxide hydrolases and dehalogenases [131]. Unlike other proteins, the core of the protein belonging to this superfamily has an α/β sheet composed of 8 β strands lined to 6 α helices [130,132].

3.3.14. Domain of Unknown Function (DUF)

Generally, every protein domain has a distinct structure and function. However, there are several domains which have no known role, and these are referred to as domains of unknown function (DUFs). Most of the time these domains were ignored as having little relevance, but now there are reports which show that many DUFs are essential, as they are crucial for protein function. Basing themselves on sets of bioinformatic analyses of several uncharacterized DUFs, Goodcare et al. speculated about probable tasks which may be related to ATP binding or transcription [133].

3.3.15. ANK Repeats

Ankyrin (ANK) like repeats mediated protein-protein interactions between diverse groups of proteins [134,135] have been reported in almost all species [136]. The ANK proteins exhibit a domain shuffling via a horizontal gene transfer [137]. A protein may have several numbers of ANK repeats per protein [134,138]. Davis et al. [138] demonstrated the association of a specific 33 residue erythrocyte ankyrin repeat with an anion exchanger. A stack of ANK repeats has a superhelical arrangement
with four consecutive repeats, and each unit contains two antiparallel helices and a beta-hairpin. ANK repeats may also occur in combinations with other types of domains [134].

3.3.16. RhoGEF Domain

The RhoGEF protein is a guanine nucleotide exchange factor (GEF) responsible for the activation of Rho family GTPases (Rho, Rac, and Cdc42) [139–141]; it controls a diverse array of cellular processes, including cellular differentiation [142], cell morphology [143], cell motility and adhesion [144], phagocytosis [145], cytokinesis [146], smooth muscle contraction [147], and the etiology of human disease such as hypertension [148] and cancer [149]. The Rho family proteins are generally found in two different conformational states, i.e., active GTP-bound and inactive GDP-bound [150]. The Rho family GTPases have a conserved domain of ~200 amino acid residues known as the RhoGEF domain or Dbl homology (DH) domain, which encodes a GEF specific to different Rho family members [151]. In addition to the RhoGEF domain, Rho family GTPases also have another functionally independent conserved domain (~100 amino acid residue), i.e., the pleckstrin homology (PH) domain, located at the C-terminus of the RhoGEF domain [152]. The C-terminal PH domain is generally involved in intracellular targeting and regulates the function of the RhoGEF domain. The RhoGEF domain has an α-helix bundle-like structure with three conserved regions, i.e., conserved region 1 (CR1), conserved region 2 (CR2) and conserved region 3 (CR3). Among these three conserved regions, CR1 and CR3 interact partly with α-6 and the DH/PH junction site, forming the Rho GTPase binding pocket.

3.3.17. PDZ Domain

PDZ domains or discs-large homologous regions (DHR) are widely spread in a wide range of membrane-bound signaling proteins from bacteria, yeasts, plants, insects and vertebrates [153,154]. The PDZ domain presents either as a single copy or as multiple copies and interacts either with the C-terminus of proteins or with internal peptide sequences [154]. Proteins with the PDZ domain are generally located at the plasma membrane, where they can directly interact with phosphatidylinositol 4, 5-bisphosphate (PIP2), as observed with the class II PDZ domain in syntentin [155]. PDZ domains (80–90 amino acids) are composed of compactly arranged six β-strands (βA-βF) and two α-helices (αA and αB) in a globular structure. PDZ domains interact with shaker-type K⁺ channels in several MAGUKs or bind similar ligands of other transmembrane receptors [154].

3.3.18. GAL4-Fungal TF MHR Domain

Gal4 is a fungal-specific positive regulator for the expression of galactose-induced genes [156]. This domain is generally located at the N-terminus of several fungal-specific transcriptional regulators and contains a binuclear Zn cluster bound by six Cys residues; additionally, it is involved in the zinc-dependent binding of DNA [157,158]. The transcriptional regulators or proteins with the GAL4-fungal TF-MHR domain are generally involved in the arginine, proline, pyrimidine, quinate, maltose and galactose metabolisms, amide and GABA catabolism and leucine biosynthesis [159].

3.4. BLUF Proteins for Optogenetic Tools

Small photoreceptors, such as sensors of blue light using flavin (BLUF), light oxygen voltage (LOV) based receptors and cryptochromes, have been identified in the genomes of different organisms, from prokaryotes to higher eukaryotes, as has been shown in the present study and in several other studies [160,161]. As evident from the present study and other in vitro experiments [161–165], one of the peculiar features of these light-sensitive motifs are their association with various effector domains, thus hinting about the wide range of novel mechanistic and functional diversity controlled by light. This aspect is yet to be studied in detail. Investigations into PixJ1 (blue and green light), RcaE (red and green), changes in light-dependent E. gracilis PAC-α activity for Drosophila, behavioral modulation and neural responses in marine gastropod Aplysia and Caenorhabditis elegans [166–171] suggest sensitivities of each receptor toward a particular light bandwidth, which can be one of the
important tools used to engineer a novel system for a broad range of physiological outputs. This study further expands the possibilities for the BLUF domain to be used as a powerful optogenetic tool for the development of novel optogenetic technologies. A vast variety of domain combinations of BLUF photoreceptors in different genomes (Figure 1) represents a promising and valuable tool to design novel photo-regulated enzymes, messengers, photo-modulation of gene expression patterns, photo-control of the virulence in pathogenic bacteria through the recombinant expression of such systems, photobehavioural responses in photobacteria, modulation of neural systems and dynamic molecular switches to regulate biological activities. BLUF domains, in combination with the EAL, GGDEF or CHD domains, could be utilized for the photo-dependent regulation of c-di-GMP and cAMP associated signaling in bacteria [64,85]. The BLUF domain associated with a B$_{12}$ binding domain has also been analyzed; however, it does not show the important amino acids that are required in order to form the binding pocket for AdoB$_{12}$. Cheng et al. [92] also reported a similar BLUF module and suggested that the associated B$_{12}$ binding domain also has a photosensory function which can regulate activity in response to light. The B$_{12}$ binding domain broadens the absorption range for the BLUF photosensor, which could be critical to several regulatory pathways [96,172]. We could also use this modular combination for the photo-dependent regulation of pathways like carotenoid synthesis or photosystem biosynthesis [92]. Another combination which could be engineered and used for an optogenetic application is the association of the BLUF domain with the DNA pol III $\gamma$ III domain. Using this modular architecture, we could regulate the actions of different components involved in DNA replication in a light-dependent manner. Furthermore, the BLUF domain is also found in association with the p450 (cytochrome p450) domain, which could be used as an optogenetic tool for the light-dependent regulation of several pathways like fatty acid metabolism and secondary metabolism. The optogenetic potential of the BLUF domain could also be acquired in the two-component regulatory system for both prokaryotes and eukaryotes. A modular architecture in which the BLUF sensor domain is associated with the Rec (receiver) domain could be exploited as a two-component regulatory system for the photo-dependent regulation of processes like bacterial chemotaxis and other regulatory pathways [118]. The BLUF domain associated with the TetR or AcrR domain may be exploited for the light-dependent regulation of pathways (as mentioned earlier) in bacteria. The BLUF domain was also analyzed with the endonuclease 3c and endonuclease_NS domains for their optogenetic potential. The BLUF domain endonuclease 3c could be engineered and used as an optogenetic tool for the light-dependent regulation of the DNA repair process. On the other hand, the BLUF domain associated with the endonuclease_NS domain could be exploited for the light-dependent regulation of processes like hydrolase activity, nucleic acid binding and metal ion binding [126]. The modular architectures comprised of BLUF associated with the AraC and abhydrolase domains could also be harnessed for the light-dependent regulation of the arabinose metabolism, as well as the diverse group of hydrolytic enzymes that include proteases, lipases, peroxidases, esterases, epoxide hydrolases, and dehalogenases, respectively [131]. BLUF in combination with the RhoGEF domain is also considered an important modular architecture, which could be used as an optogenetic tool for the light-dependent regulation of a diverse array of cellular processes [142–146,148,149].

**4. Conclusions**

Applications of the photoreceptors in order to quickly control molecular machines and, in turn, biological systems and processes, present the scientific community with an exciting opportunity with several possibilities, approaches, along with their limitations as well. Recently, several such approaches have been reported, including photoswitches, UV photo-reactivation and deactivation, spectral tuning, and photocaging [25]. However, a detailed understanding of electron transfer mechanisms, transient state intermediates, and amino acid patterns will allow the development of more precise
Supplementary Materials: The following are available online at http://www.mdpi.com/2076-3417/9/18/3924/s1,
Figure S1: Multiple sequence alignment of the BLUF coupled EAL domain. Sequence representing the EAL domain from the BlrP1 protein was used as the template for the sequence alignment analysis. Amino acid residues in solid box are the conserved residues involved in the formation of the EAL active site, Figure S2: Multiple sequence alignment of the BLUF coupled CHD domain. Sequence representing the CHD domain from the bPAC protein was used as template for the sequence alignment analysis. Amino acid residues in solid box are the conserved residues involved in the formation of the nucleotide binding site, Figure S3: Multiple sequence alignment of the BLUF coupled PAS domain. Sequence representing the PAS domain from the photactivated yellow protein (PYP) from Halorhodospira halophila was used as template for the sequence alignment analysis. Amino acid residues in solid box are the conserved residues involved in the formation of the PA6 active site, Figure S4: Multiple sequence alignment of the BLUF coupled vitamin B12 binding domain. Sequence representing the B12 binding domain from the CarHI and AerR proteins was used as template for the sequence alignment analysis. The conserved amino acids (Trp 131, Val138, Glu141 and His142) essential for forming the binding pocket for the substrate, i.e., AdoB12, is not found in the aligned portion of the BLUF coupled vitamin B12 binding domain, Figure S5: Multiple sequence alignment of the BLUF coupled DNA pol III γ III domain. The sequence of the well characterized truncated (1-373 amino acids) DNA polymerase III subunit gamma/tau (WP_113440333.1) from E. coli was used as template for the sequence alignment analysis. Amino acid residues in solid box represent the important residues crucial for the enzyme activity, Figure S6: Multiple sequence alignment of the BLUF coupled p450 and the well characterized CYP protein from Bacillus subtilis (used as template). Amino acid residues in solid box representing the conserved (Arg) and altered (Pro to Ser) amino acid residues essential for the substrate binding, Figure S7: Protein-Protein interaction network depicting interacting partners of the selected effector domain (EAL, CHD, PAS, B12, DNA POL III γ III and p450) of the BLUF modular proteins. Protein highlighted in yellow is the query protein. Protein-protein interaction analysis was performed by using String version 11 (https://string-db.org/). Details of the query proteins, domains, along with the annotations are given in Table S1, Table S1: Output showing the details of query proteins, domains, interacting proteins and annotated functions.

Author Contributions: M.S.K. had analysed all the sequences in details and wrote the manuscript. R.S. had done preliminary analysis of the BLUF coupled proteins. S.K.V. had conceptualized protein-protein interaction of the BLUF coupled domains and wrote manuscript. S.K.S. had done detailed structure-function analysis of the BLUF and its modular domains and wrote relevant part of the paper. S.K. had conceptualized, outlined and wrote the manuscript.
**Funding:** M.S.K. research associateship was supported from SR/NM/NB-1087/2017(G)-JNU. S.K. was supported by the SR/NM/NB-1087/2017(G)-JNU. S.K.S. was supported by the research endowment fund from MUJ.

**Conflicts of Interest:** The authors have no conflict of interest.

**References**


102. Cotton, R.G.; Gibson, F. The biosynthesis of phenylalanine and tyrosine; enzymes converting chorismic acid into prephenic acid and their relationships to prephenate dehydratase and prephenate dehydrogenase. *Biochim. Biophys. Acta* 1965, 100, 76–88. [CrossRef]


108. Li, M.; Gu, R.; Su, C.C.; Routh, M.D.; Harris, K.C.; Jewell, E.S.; McDermott, G.; Dixon, N.E. The unstructured C-terminus of the g subunit of *Escherichia coli* DNA polymerase III holoenzyme is the site of interaction with the a subunit. *Nucleic Acid Res.* **2007**, *35*, 2813–2824. [CrossRef] [PubMed]


127. Miller, M.D.; Tanner, J.; Alpaugh, M.; Benedik, M.J.; Krause, K.L. 2. 1 Å structure of *Serratia* endonuclease suggests a mechanism for binding to double-stranded DNA. *Nat. Struct. Mol. Biol.* 1994, 1, 461. [CrossRef]


138. Davis, L.H.; Otto, E.; Bennett, V. Specific 33-residue repeat (s) of erythrocyte ankyrin associate with the anion horizo? [CrossRef] [PubMed]


© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).