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

Manish Singh Kaushik, Ramandeep Sharma, Sindhu KandothVeetil, Sandeep Kumar Srivastava and Suneel Kateriya

Table S1. String analysis [1] output showing the details of query proteins, domains, interacting proteins and annotated functions.

S. No.	Query protein	Domain	Interacting Partner	Annotation
1.	JD73_24940	EAL (Diguanylate cyclase)	JD73_03740	C-di-GMP phosphodiesterase
			YeaP	Diguanylate cyclase
			JD73_23680	Diguanylate cyclase
			JD73_23675	Diguanylate cyclase
			YdaM	Diguanylate cyclase
			AriR	Regulator of acid resistance
			YcgZ	Two-component-system connector protein
			YcgE	HTH-type transcriptional regulator, MerR domain protein
			JD73_25605	Regulatory protein MerR
			GJ12_01945	Transcriptional regulator
2.	AMSG_04679	CHD (class III nucleotydyl cyclase)	AMSG_00147	Phosphodiesterase
			AMSG_00905	Phosphodiesterase
			AMSG_01576	Uncharacterized protein
			AMSG_01591	Adenylyl cyclase-associated protein belongs to the CAP family
			AMSG_04591	DNA-directed RNA polymerase subunit beta
			AMSG_08774	Uncharacterized protein
			AMSG_08967	cGMP-dependent 3',5'-cGMP phosphodiesterase A
			AMSG_09378	Adenylate/guanylate cyclase with GAF and PAS/PAC sensor
			AMSG_10048	3,4-dihydroxy-2-butanone 4-phosphate synthase
			AMSG_11978	DNA helicase
3.	Hhal_1818	PAS (blue light sensor)	Hhal_0366	Multi-sensor hybrid histidine kinase
			Hhal_0474	CheA signal transduction histidine kinase
			Hhal_0522	Putative CheW protein
			Hhal_0934	CheA signal transduction histidine kinase
			Hhal_1716	CheBmethylesterase; MCP methyltransferase, CheR-type
			Hhal_1819	4-coumarate-CoA ligase
			Hhal_1820	Phenylalanine/histidine ammonia-lyase
			Hhal_2150	Multi-sensor hybrid histidine kinase
			Hhal_2151	Hpt sensor hybrid histidine kinase
			Hhal_2167	CheA signal transduction histidine kinase
4.	Rsph17025_3045	Vitamin B_{12} binding	Rsph17025_0524	Prephenate dehydratase
			tyrS	Tyrosyl-tRNA synthetase
			Rsph17025_0963	Hypothetical protein
			Rsph17025_1890	Hypothetical protein
			Rsph17025_2731	SARP family transcriptional regulator
			Rsph17025_2732	PA-phosphatase-like phosphoesterase
			mutL	DNA mismatch repair protein
			Rsph17025_2994	TonB family protein
			Rsph17025_3046	Hypothetical protein
			Rsph17025_3047	Cobalamin (vitamin B12) biosynthesis CbiX protein

5.	DnaX	DNA PolIII_γIII (DNA polymerase III, subunits gamma and tau)	dnaE	DNA polymerase III subunit α
			dnaQ	DNA polymerase III subunit ε
			dnaN	β sliding clamp
			dnaB	Replicative DNA helicase
			recR	Recombination protein
			holA	DNA polymerase III subunit δ
			holB	DNA polymerase III subunit δ
			holC	DNA polymerase III subunit chi
			holD	DNA polymerase III subunit psi
			polA	DNA polymerase I
6.	CypC	p450 (monooxygenase)	cypD	Bifunctional cytochrome P450/NADPH-P450 reductase 1
			hemE	Uroporphyrinogen decarboxylase
			yitS	DegV domain-containing protein involved in lipid transport
			yjiB	Putative cytochrome P450
			pksJ	Intermediate polyketide synthase involved in secondary metabolism
			pksM	Intermediate polyketide synthase involved in secondary metabolism
			cypB	Bifunctional cytochrome P450/NADPH-P450 reductase 2
			yrzI	Uncharacterized protein
			bioI	Biotin biosynthesis cytochrome P450
			cypX	Pulcherriminic acid synthase

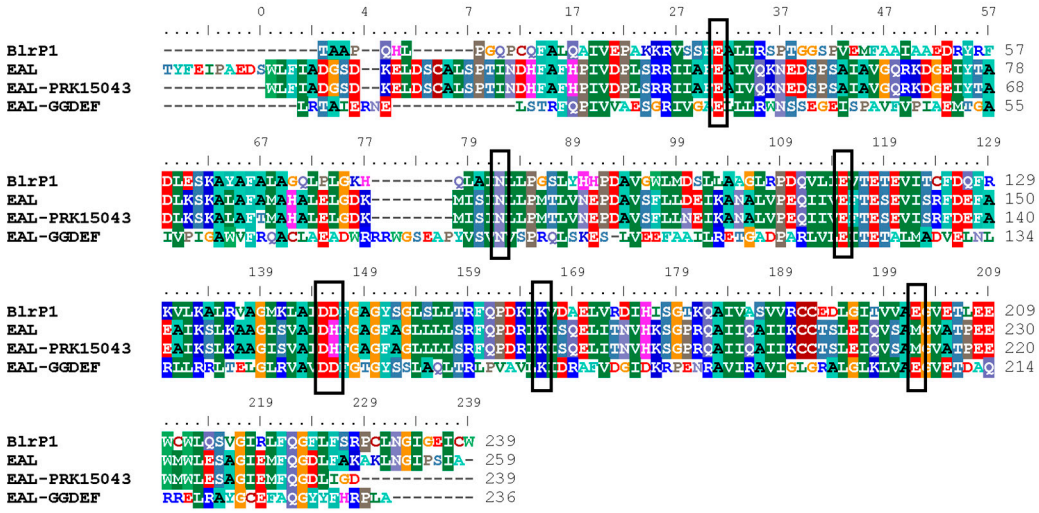


Figure S1. Multiple sequence alignment of the BLUF coupled EAL domain using BioEdit tool [2]. Sequence representing EAL domain from BlrP1 protein was retrieved from National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) and used as template for the sequence alignment analysis. Amino acid residues in solid box are the conserved residues involved in the formation of EAL active site.

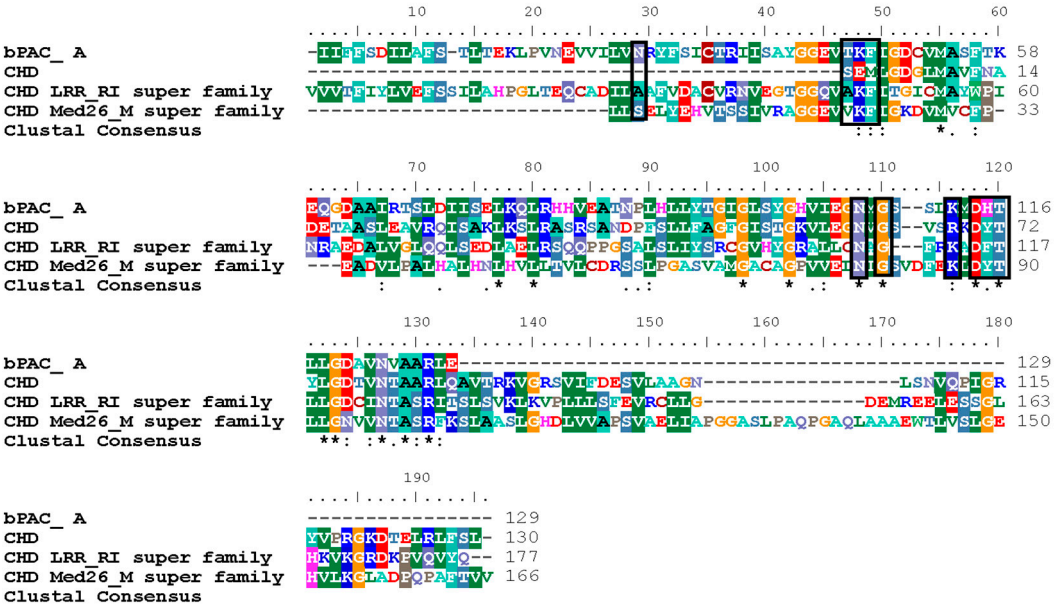


Figure S2. Multiple sequence alignment of the BLUF coupled CHD domain using BioEdit tool [2]. Sequence representing CHD domain from bPAC protein was used as template for the sequence alignment analysis. Amino acid residues in solid box are the conserved residues involved in the formation of nucleotide binding site.

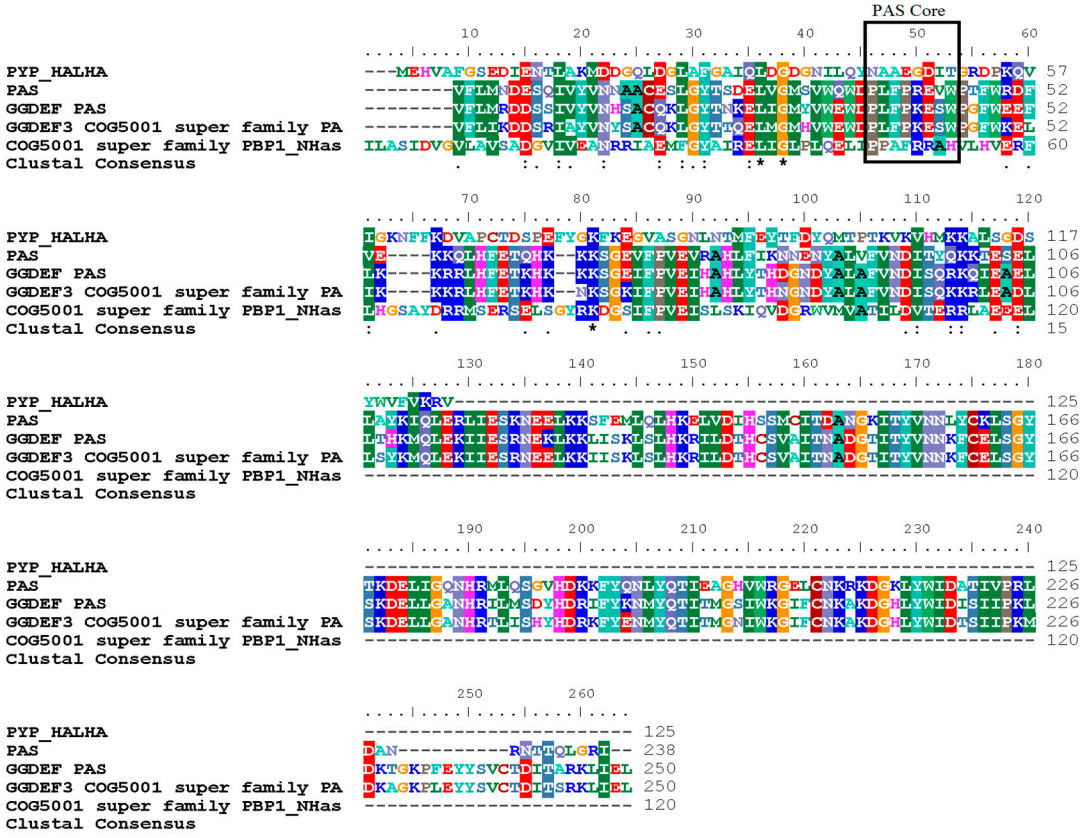


Figure S3. Multiple sequence alignment of the BLUF coupled PAS domain using BioEdit tool [2]. Sequence representing PAS domain from photoactivated yellow protein (PYP) from *Halorhodospira halophila* was retrieved from National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) and used as template for the sequence alignment analysis. Amino

acid residues in solid box are representing the PAS core motif responsible for generation and propagation of signal to the adjoining effector domain.

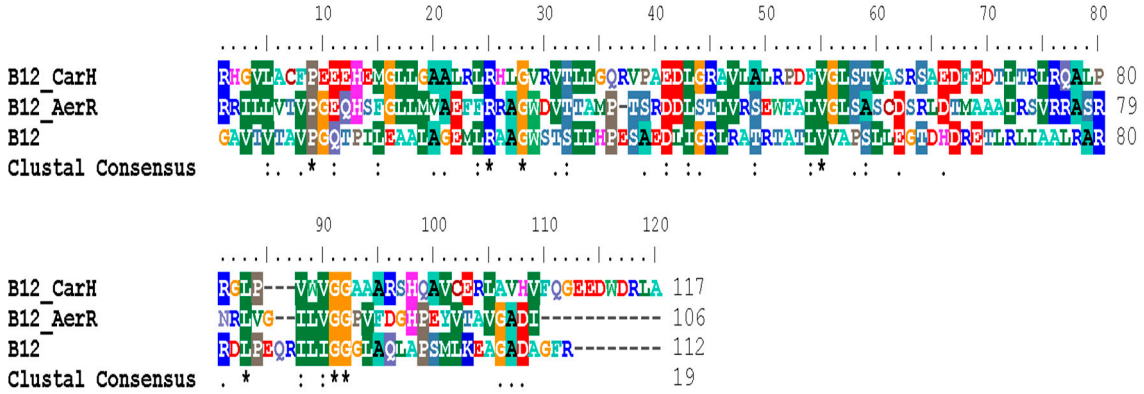


Figure S4. Multiple sequence alignment of the BLUF coupled vitamin B_{12} binding domain using BioEdit tool [2]. Sequence representing B_{12} binding domain from CarH and AerR proteins were retrieved from National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) and used as template for the sequence alignment analysis. The conserved amino acids (Trp 131, Val138, Glu141 and His142) essential for forming the binding pocket for substrate i.e. AdoB_{12}, is not found in the aligned portion of BLUF coupled vitamin B_{12} binding domain.

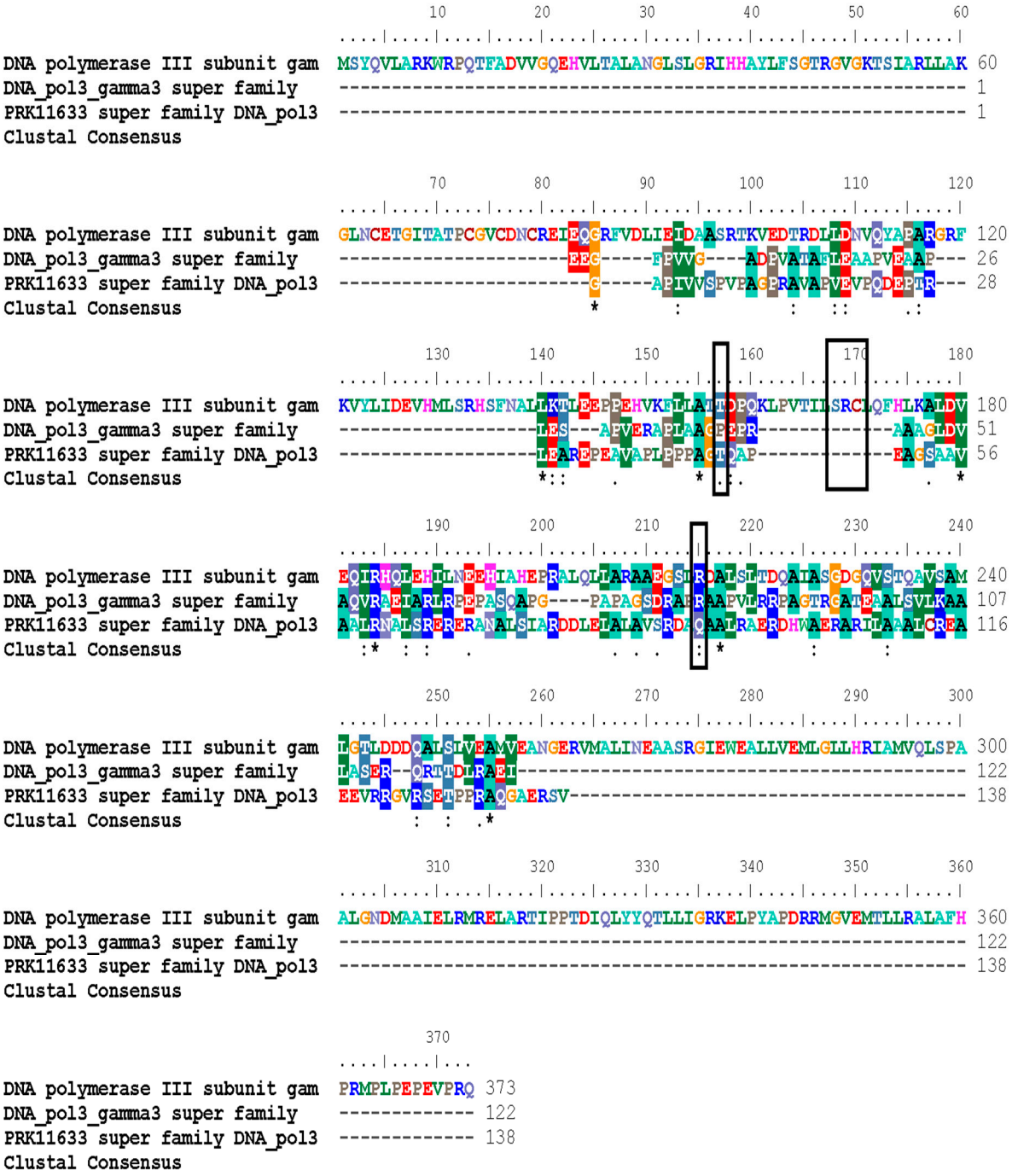


Figure S5. Multiple sequence alignment of the BLUF coupled DNA pol III γ III domain using BioEdit tool [2]. The sequence of the well characterized truncated (1-373 amino acids) DNA polymerase III subunit gamma/tau (WP_113440333.1) from *E. coli* was retrieved from National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) and used as template for the sequence alignment analysis. Amino acid residues in solid box represents the important residues crucial for the enzyme activity.

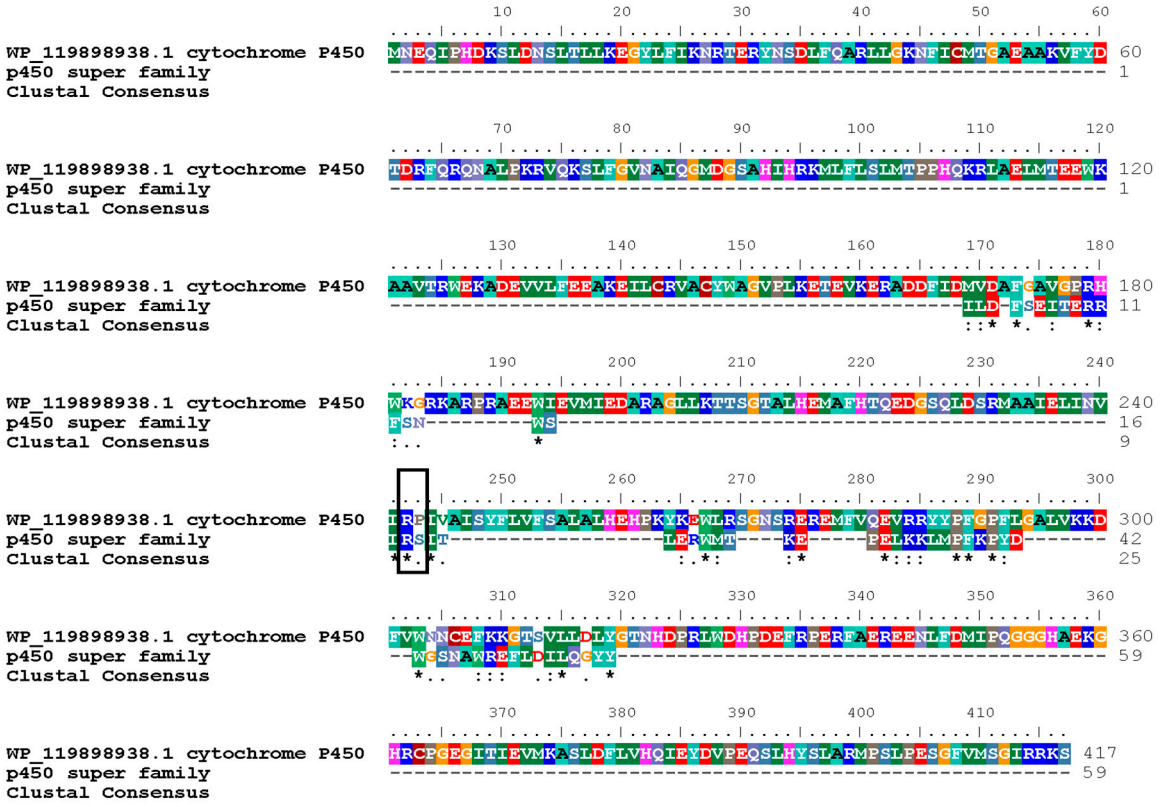
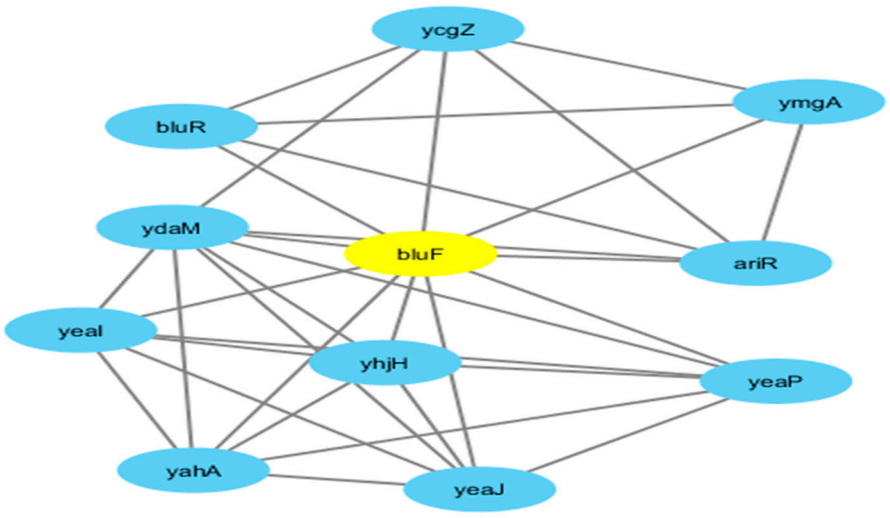
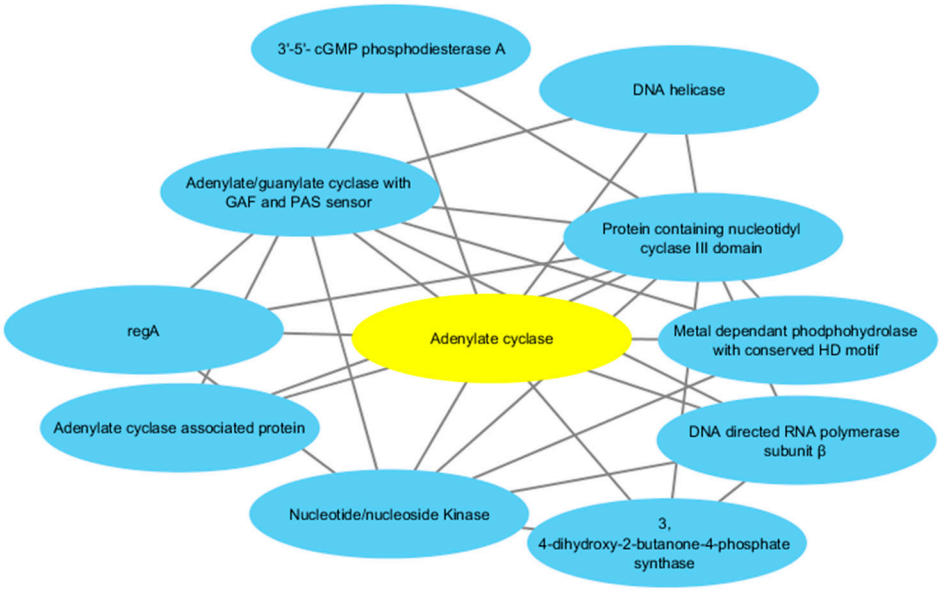


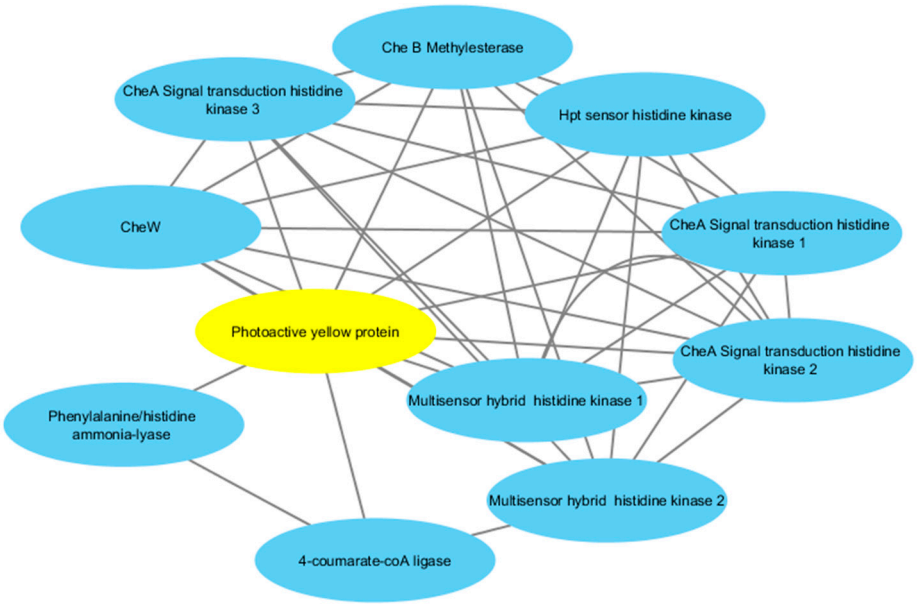
Figure S6. Multiple sequence alignment of the BLUF coupled p450 and well characterized CYP protein from *Bacillus subtilis* (retrieved from National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) and used as template) using BioEdit tool [2]. Amino acid residues in solid box representing the conserved (Arg) and altered (Pro to Ser) amino acid residues essential for the substrate binding.



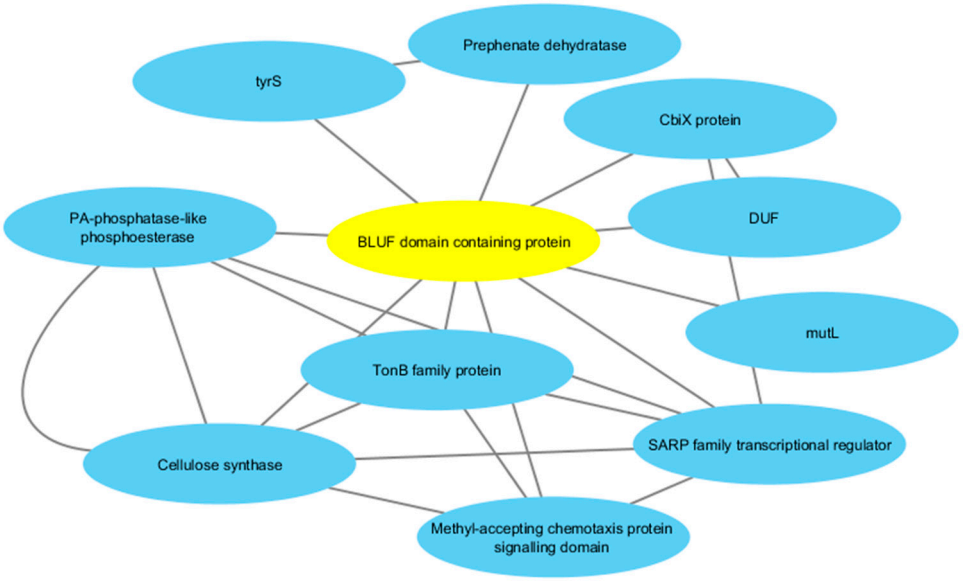
(**a**)


3'-5'- cGMP phosphodiesterase A
DNA helicase
Adenylate/guanylate cyclase with GAF and PAS sensor
Protein containing nucleotidyl cyclase III domain
regA
Adenylate cyclase
Metal dependant phodphohydrolase with conserved HD motif
Adenylate cyclase associated protein
DNA directed RNA polymerase subunit β
Nucleotide/nucleoside Kinase
3, 4-dihydroxy-2-butanone-4-phosphate synthase

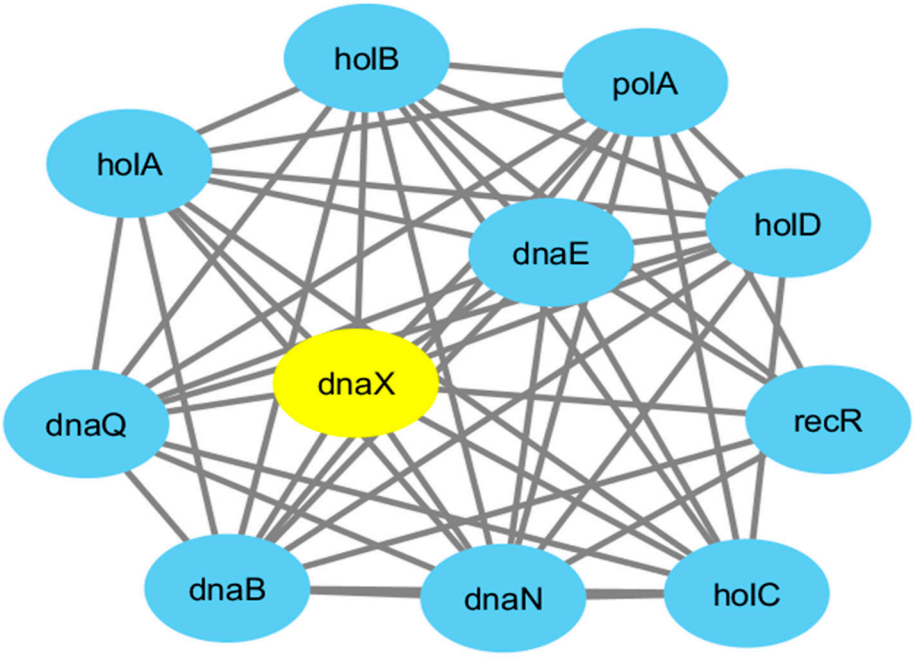
(**b**)


Che B Methylesterase
CheA Signal transduction histidine kinase 3
Hpt sensor histidine kinase
CheW
CheA Signal transduction histidine kinase 1
Photoactive yellow protein
CheA Signal transduction histidine kinase 2
Multisensor hybrid histidine kinase 1
Phenylalanine/histidine ammonia-lyase
Multisensor hybrid histidine kinase 2
4-coumarate-coA ligase

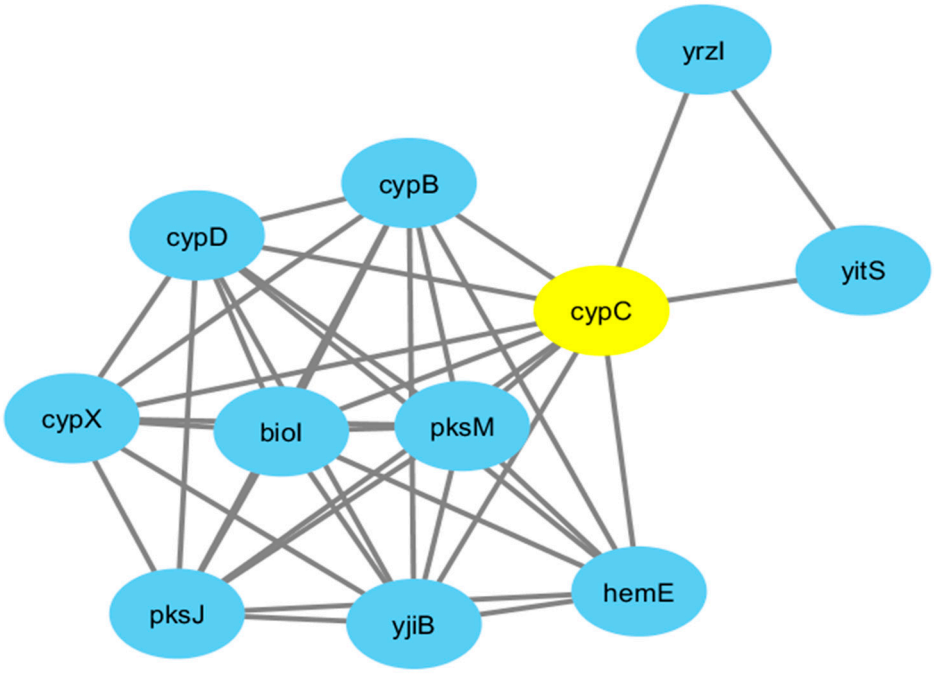
(**c**)


Prephenate dehydratase
tyrS
CbiX protein
PA-phosphatase-like phosphoesterase
BLUF domain containing protein
DUF
mutL
TonB family protein
SARP family transcriptional regulator
Cellulose synthase
Methyl-accepting chemotaxis protein signalling domain

(**d**)


holB
polA
holA
holD
dnaE
dnaX
dnaQ
recR
dnaB
dnaN
holC

(**e**)



(**f**)

Figure S7. Protein-Protein interaction **n**etwork depicting interacting partners of the selected effector domains (EAL, CHD, PAS, B_{12}, DNA POL III γ III and p450) of the BLUF modular proteins. Protein-protein interaction analysis was performed using String version 11 (https://string-db.org/) [2] and further modified using CytoScape [3]. Protein highlighted in yellow is the query protein. (**a**) The protein-protein interaction analysis of query protein sequence containing the EAL output domain revealed several interacting partners belongs to either EAL domain-containing PDEs or GGDEF domain-containing DGCs, (**b**) Protein-protein interaction analysis CHD domain containing protein showing the possible interacting partners, which range from phosphodiesterases, the RNA polymerase subunit β, and the DNA helicase to another adenylate cyclase/guanylate cyclase – associated with the GAF and PAS/PAC sensor,(**c**) Protein-protein interaction analysis PAS domain containing protein showing the possible interactions with CheW, CheA signal transduction histidine Kinase 1, 2 and 3, multisensor histidine Kinase 1 and 2, CheB methyltransferases, Hpt sensor histidine kinase,(**d**) The protein-protein interaction analysis for query protein containing the B_{12} domain showed several interacting partners involved in the regulation of different signaling pathways like prephenate dehydratase enzyme, tyrosyl-tRNA synthetase (TyrS) and a DNA mismatch repair protein, MutL, ,(**e**) The protein-protein interaction analysis for DNA pol III γ III (dnaX) revealed the interacting partners are the components involved in the regulation of DNA replication, i.e. 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),(**f**) The protein-protein interaction analysis for protein containing p450 domain showed interacting partners belonged to the fatty acid metabolism (CypB, CypC, CypD, and YitS) and to the secondary metabolism (PksJ and PksM). Details of the query proteins, domains along with the annotations are given in Table S1.

References

1. Szklarczyk, D.; Morris, J.H.; Cook, H.; Kuhn, M.; Wyder, S.; Simonovic, M.; Santos, A.; Doncheva, N.T.; Roth, A.; Bork, P.; et al. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **2016**, *45*, 362–368.
2. Hall, A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for window 95/98/NT. *Nucliec Acid Res.* **1999**, *41*, 95–98.

3. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. CytoScape: a software environment for integral model of biomolecular interaction network. *Genome Res.* **2003**, *13,* 2498–2504.