BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 8(7)

Homology modeling, docking studies and functional analysis of various azoreductase accessory interacting proteins of *Nostoc sp.*PCC7120

Priyadarshini Devi Philem & Samrat Adhikari*

Bioinformatics Infrastructure Facility, Department of Biotechnology, St Edmund's College, Shillong- 793003, Meghalaya, India; Samrat Adhikari - Email: stedmundc.btisnet@nic.in, naimamode@gmail.com; Phone- 91-0364-2220808; 91-9862041757; *Corresponding author

Received March 28, 2012; Accepted April 04, 2012; Published April 13, 2012

Abstract:
Azo dyes have become a threat to public health because of its toxicity and carcinogenicity. Azoreductase enzyme plays a pivotal role in the degradation of azodyes released by industrial effluents and other resources. The degradation pathway has to be studied in detail for increasing the activity of azoreductase and for better degradation of azo dyes. But the data available on cyanobacterial azoreductase enzyme and its degradation pathway are still very less. Therefore the present work explored the azoreductase pathway of the cyanobacterium *Nostoc sp.* PCC7120 for better understanding of the degradation pathway and the other accessory interacting proteins involved. The accessory interacting proteins of azoreductase from cyanobacterium *Nostoc sp.* PCC7120 were obtained from STRING database. The proteins do not have a comprehensive three dimensional structure and are hypothetical. The secondary structure and functional analysis indicated that the proteins are all soluble proteins, without disulphide bonds and have alpha helices only. The structural prediction and docking study showed that alr2106, alr1063 and alr2326 have best docking result which tally with the STRING database confidence score and thus these proteins could possibly enhance the azoreductase activity and better dye degradation. These results will pave way for further increase in azoreductase activity and for better understanding of the dye degradation pathway.

Keywords: *Nostoc sp. PCC7120*, azoreductase accessory proteins, molecular modeling tools, docking, cyanobacterium.

Background:
Azo dyes, aromatic moieties linked together by azo (–N=N–) chromophores represent the largest class of dyes which have wide applications in textiles, leather, plastics, cosmetics, printing, paper making, and cosmetic industries for their versatility. These residual dyes in industrial effluents are a threat to public health because of its high toxicity and carcinogenicity **[1, 2].** Many microorganisms including cyanobacteria could transform these azo dyes into colorless products, which are initiated by an enzymatic step involving cleavage of azo linkages with the aid of an azoreductase enzyme and an electron donor **[3 - 8].** Azoreductase enzyme of many microbes confers potential advantages in the bioremediation of dyes contaminated areas **[9].** The first step in the biodegradation of azo compounds is the reduction to the corresponding amines, a reaction catalyzed by azoreductase resulting to aromatic amines which are further degraded aerobically **[10]**. Similarly azoreductase activity of cyanobacterial species which are known for their ubiquitous occurrence in nature and the activity of this enzyme in response to various mono and diazo compounds have also been reported **[11, 12],** but a detailed analysis on the degradation pathway and the involvement of other enzymes or proteins in general has not been addressed so far. There is thus an increasing demand on understanding the function and relationships of proteins in the degradation pathway. Therefore the present work signifies an

attempt to study the cascade of reaction occurring during azodye degradation by the azoreductase enzyme in cyanobacteria using systems biology approach. Most proteins attain their biological functions through specific interactions with other proteins. Thus, the study of protein-protein interactions and the interfaces that mediate these interactions is of prime importance for the understanding of biological function **[13]**. It is important to understand the enzyme degradation pathway, its components and the interaction taking place in the pathway if we plan to regulate the pathway for increasing the azoreductase activity and better degradation of the azodyes. Systems biology being a holistic and a pivotal approach which involves various molecular modeling, metabolic pathways analysis, and regulatory and signal transduction networks for understanding better cellular behavior. There are also various levels of abstraction at which variety of techniques that has been employed based on the quality and quantity of data available **[14].** An excellent application of systems biology in metabolic engineering with commercial potential has been illustrated for improving lysine production using *E. coli* to discover putative genes impacting network properties and cellular phenotype **[15].** Hence in this context to the present study, the azoreductase enzyme (azoR) from the cyanobacterium *Nostoc sp*. PCC7120 has been accounted for exhibiting its interaction with its accessory proteins obtained from the STRING database **[16].** The interacting proteins are further studied for its structural, functional, physiochemical properties and its localization using bioinformatics tools to have detailed information on each of the components of the azodye degradation pathway. This approach might be proven to be a novel one for the identification of target genes, which could increase the azoreductase activity for accelerating azodyes degradation in context to cyanobacteria.

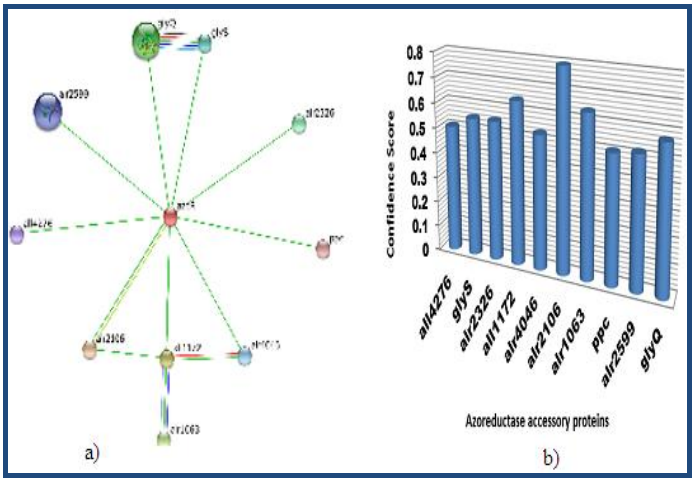

Figure 1: (a) Showing the confidence interval map of the seven azoR interacting accessory proteins that plays a pivotal role in azodye degradation pathway**; (b)** Graph representing the confidence score of the accessory proteins of azoreductase (azoR) enzyme. The scores are in the category of Highest - 0.9; High - 0.7; Medium - 0.400 and low - 0.15.

Methodology:

Sequence Retrieval

The sequence of the azoreductase enzyme was retrieved from National Centre for Biotechnology Information protein database at http://www.ncbi.nlm.nih.gov/protein having accession No: Q8YV76 which was isolated from *Nostoc sp*. PCC7120 **[17]**. Confidence interval map of azoreductase accessory proteins were analyzed from STRING database at http://string-db.org/ **[16]** and availability for authentic structures in Protein data bank was checked comparatively in NCBI Entrez, PDB and SWISSPROT Databases. The protein sequences of alr2599, all4276, alr2326, alr2106, all1172, alr1063 and alr4046 with undefined function and structure were retrieved from the STRING database. This search tool identifies interaction partners according to a variety of criteria including co localization in one or more genomes (e.g., operon structures), co-occurrence across genomes (phylogenetic profiling), and, where available, correlated expression data and / or literature citations.

Domain Analysis

The sequences of the accessory proteins were analyzed for domain architecture using the NCBI domain database CDART at www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi **[18]**.

Physiochemical Characterization

The physiochemical characterization like theoretical isoelectric point (pI), molecular wt, no. of positively and negatively charged residues, extinction coefficient, instability index, aliphatic index and Gram Average Hydropathicity (GRAVY) were computed using the Expasy Protparam server at http://web.expasy.org/protparam/ **[19]**.

Secondary structure prediction and functional analysis

The accessory proteins were characterized for its nature of solubility or presence of transmembrane regions using membrane protein prediction system, SOSUI at http://bp.nuap.nagoya-u.ac.jp/sosui/ **[20]** and disulphide linkages were computed by DISULFIND server at http://disulfind.dsi.unifi.it/ **[21]**. The Secondary structure was predicted using PSI-PRED server at http://bioinf.cs.ucl.ac.uk/psipred/ **[22].**

Model building and evaluation

The homology modeling for the seven accessory proteins were performed using CPHmodels server at www.cbs.dtu.dk/services/CPHmodels/ **[2**3] with templates 2hzt, 1tq5, 1yyv, 1lfp, 2ocs, 3acl and 3ppu respectively generated with highest similarity percentage. Evaluation of the 3D structures modeled was carried out using MOLPROBITY server at http://molprobity.biochem.duke.edu/ **[24].** The constructed models were energy minimized by QMEAN server at http://swissmodel.expasy.org/qmean/cgi/index.cgi **[25, 26].** The overall stereo chemical properties of the protein were analyzed using QMEAN server and viewed under RASMOL at www.openrasmol.org **[27].**

Analysis of Ligand binding sites and pockets

Ligand binding sites were predicted by QSITE FINDER at http://www.modelling.leeds.ac.uk/qsitefinder/ **[28].**

Docking studies

Docking study of the azoreductase enzyme and the accessory proteins were carried out using PATCHDOCK server at http://bioinfo3d.cs.tau.ac.il/PatchDock/ **[29]** and energy minimization was performed before and after docking using QMEAN server.

Discussion:

The confidence interval map of azoreductase interacting proteins and the confidence scores of each protein are depicted in **(Figure 1).** There are basically 10 accessory interacting proteins which are involved with azoreductase enzyme (azoR) activity as revealed from the STRING database. The confidence scores of all the proteins are evaluated to be good and all the interacting proteins are neighboring proteins. The interacting proteins chosen for analysis are alr2599, all4276, alr2326, alr2106, all1172, alr1063 which are hypothetical and do not have a comprehensive three dimensional structure. So functional analysis is carried out using CDART for domain prediction and resulted with important domains responsible for transcription and xenobiotic degradation **Table 1 (see supplementary material**). The protein alr2106 constitutes of domain which belonged to the family of DNA-binding protein that acts as a positive regulator. The physicochemical parameters are computed using Protparam and are compiled in **Table 2 (see supplementary material)**. Analysis of all the protein sequences with DISULFIND showed the absence of disulfide bond in all the proteins. Proteins with disulfide bonds are extracellular proteins and hence these interacting proteins are intracellular **[30].** Analysis with SOSUI revealed that the interacting proteins are all soluble proteins and are present in the cytoplasm and do not contain any signal peptides.

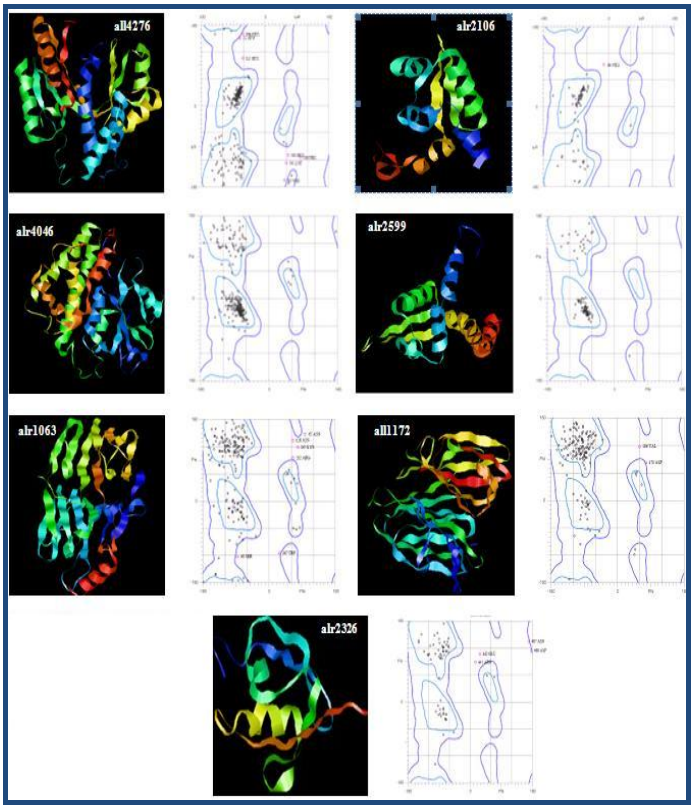

Figure 2: Homology modeled structure and the corresponding Ramamchandran plot analysis of the modeled structure.

Hence these proteins do not belong to the family of transmembrane proteins. Secondary structure prediction using PSIPRED revealed the average number of amino acids favoring alpha helix in all the proteins comparatively higher than any other secondary structure which implies the fact that intracellular proteins contain amino acids favoring alpha helix than any other secondary structure **[31]**. The result obtained from SOSUI, DISULFIND and PSIPRED revealed the intracellular localization of the accessory proteins and hence give an insight into the site of azo dyes degradation. Homology modeling of the seven accessory proteins was further carried out using CPHmodels and the resulting structures along with the Ramachandran plots are illustrated in **(Figure 2).** 3D model quality estimation for the hypothetical proteins are carried out using QMEAN Z-scores. Binding pockets prediction of azoR revealed many binding sites indicating that there is more than one important site, giving insights to competitive inhibition with the active site or the presence of multiple substrate binding sites. The docking study showed **(Figure 3, Table 3 supplementary material),** the protein alr2106 has the best docking score with azoR which also have the best confidence score obtained from STRING database. It is than followed by proteins alr2326 and alr1063 which are involved in substrate specification and detoxification. Amongst all the accessory proteins alr2106 showed the best confidence score, best docking score and the CDART result showed that it consists of DNA-binding domain that acts as positive regulator. It is followed by other proteins such as alr1063 and alr2326 which are involved in substrate specification and xenobiotic degradation. This establishes a relationship between azoR and the seven hypothetical azoreductase accessory proteins of the azoreductase (azoR) enzyme pathway.

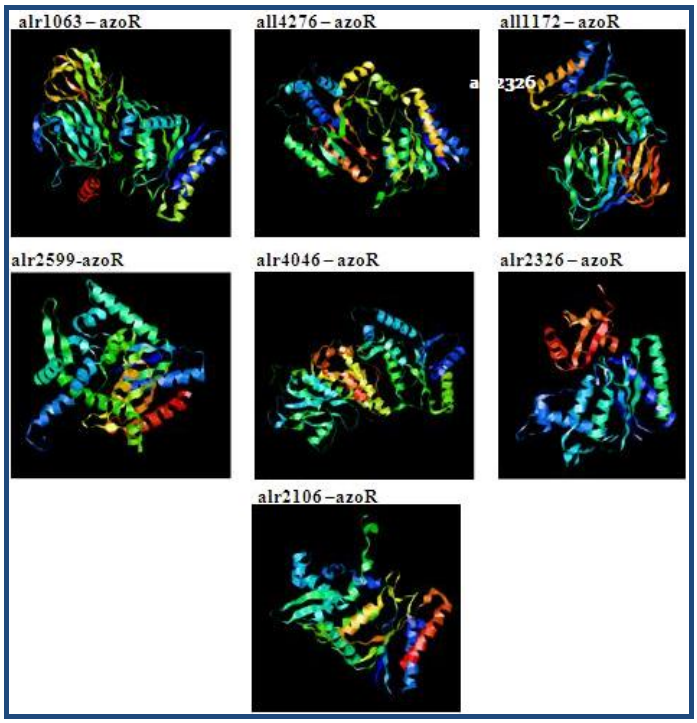

Figure 3: Docking result of the seven accessory proteins with azoreductase enzyme. The best docking result was obtained for alr2106, alr2326, alr1063 with the best global energy.

Conclusion:

Knowledge of important accessory proteins in such pathways enables identification of appropriate targets and co-targets. Here from this study we can propose transcription regulator, alr2106, and other important proteins, alr1063, alr2326 involved in substrate specification and xenobiotic degradation which can be considered for targeting the azodye degradation pathway.

Identifying the location of ligand binding sites will help us in comparing the different functional sites of azoreductase and the aspect of broad substrate specificity. Secondary structure prediction and functional analysis will give an insight to the location of these proteins along with the site of dye degradation. This information will further help us in understanding the azodye take up by the microbes and how it has been degraded.

Acknowledgements:
We take this opportunity to acknowledge the funding received from the Department of Biotechnology, Govt. of India for setting up BIF facility under BTISNET programme in Department of Biotechnology, St. Edmund's College, Shillong. We also express our heartfelt gratitude to Dr Sylvanus Lamare, Principal, St. Edmund's College, Shillong for his support throughout the work.

References:

[1] Bisschops I & Spanjers H, *Environ Technol.* 2003 **24**: 1399 [PMID: 14733393]
[2] Weisburger JH. *Mutat Res.* 2002 **507**: 9[PMID: 12351140]
[3] Asgher M. *et al*. *Biodegradation*. 2007 **18**: 311[PMID: 17004031]
[4] Hong YG & Gu JD, *Appl Microbiol Biotechnol*. 2010 **88**: 637 [PMID: 20706834]
[5] Delee WJ et al. *Chem Technol Biotechnol*.1998 **73**: 323
[6] Levine WG. *Drug Metab Rev*. 1991 **23**: 253[PMID: 1935573]
[7] Walker R, *Food Cosmet Toxicol*. 1970 **8**: 659[PMID: 5500003]
[8] Yeh MS *et al. Biotechnol Prog*. 2005. **21**: 1329 [PMID: 16080719]
[9] Kolekar YM *et al. Bioresour Technol.* 2012 **104**: 818 [PMID: 22153293]
[10] Sarayu K & Sandhya S, *Appl Biochem Biotechnol*. 2010 **160**: 1241[PMID: 19277481]
[11] Jadhav SU *et al. J Microbiol Biotechnol*. 2008 **19**: 409
[12] Omar HH & Pak J, *Biol Sci*. 2008 **11:** 1310[PMID: 18817261]
[13] Talavera D *et al. PLoS One*. 2011 6: e21053 [PMID: 21738603]
[14] Papin JA *et al*. *Nat Rev Mol Cell Biol.* 2005 **6**: 99 [PMID: 15654321]
[15] Alper H *et* al. *Metab Eng.* 2005 **7:** 155
[16] Szklarczyk D *et al*. *Nucleic Acids Res.* 2011 **39**: D561
[17] Kaneko T *et al. DNA Res*. 2001 **8**: 205 [PMID: 11759840]
[18] Geer LY *et al. Genome Res.* 2002 **12**: 1619 [PMID: 12368255]
[19] Gasteiger E *et al*. *Methods Mol Biol*. 1999 **112**: 531 [PMID: 10027275]
[20] Hirokawa T *et al*. *Bioinformatics*. 1998 **14**: 378 [PMID: 9632836]
[21] Ceroni A *et al*. *Nucleic Acids Res*. 2006 **3**: W177
[22] McGuffin LJ *et al*. *Bioinformatics.* 2000 **16**: 404 [PMID: 10869041]
[23] Nielsen M *et al*. *Nucleic Acids Res.* 2010 **38** :W576
[24] Chen VB *et al*. *Acta Crystallogr D Biol Crystallogr*. 2010 **66**: 12 [PMID: 20057044]
[25] Benkert P *et al*. *Nucleic Acids Res*. 2009 **37**: W510
[26] Benkert P *et al*. *Proteins*. 2008 **71**: 261[PMID: 17932912]
[27] Sayle RA & Milner-White EJ, *Trends Biochem Sci*. 1995 **20**: 374 [PMID: 7482707]
[28] Lovell SC *et al*. *Proteins*. 2003 **50**: 437 [PMID: 12557186]
[29] Laurie AT & Jackson RM, *Bioinformatics.* 2005 **21**: 1908 [PMID: 15701681]
[30] Schneidman-Duhovny D *et al*. *Nucleic Acids Res*. 2005 **33:** W363
[31] Nishikawa K *et al*. *J Biochem*. 1983 **94**: 997 [PMID: 6643433]

Edited by P Kangueane
Citation: **Philem** & **Adhikari,** Bioinformation 8(7): 296-300 (2012)

BIOINFORMATION

open access

Supplementary material:

Table1: Azoreductase accessory proteins obtained from STRING database are subjected to functional analysis. Domain analysis of these proteins indicated the presence of important functional domains such as domains involved in transcription regulation, substrate specification and xenobiotic degradation. Here the accession number from STRING database, the number of amino acid composition for each protein and the description of the domain are listed for each interacting proteins.

Azoreductase accessory interactive proteins	Accession number from STRING database	Length (Total no of amino acids)	Description of the Hypothetical seven azoreductase accessory proteins
alr2599	alr2599	131	Consists of winged helix-turn-helix (WHTH) DNA-binding domain of the GntR family of transcriptional regulators
all4276	all4276	252	Belongs to the family of transcriptional activator of cytochrome C oxidation (TACO1)
alr2326	alr2326	110	Consists of domain such as PDZ domain in association with glycyl aminopeptidase which are involved in peptide binding grove and which have got preference for substrates with an terminal glycine or alanine
alr2106	alr2106	238	Consists of domain which belonged to the family of DNA-binding protein that acts as a positive regulator of the formaldehyde-inducible hxIAB operon in *Bacillus subtilis*
all1172	all1172	311	Consists of domain belonging to the family of pirin proteins. It also consists of other domains such as cupin domain and pirin C-terminal cupin domain
alr1063	alr1063	335	Consists of domain belonging to the family GST_C_family Superfamily. A large, diverse group of cytosolic dimeric proteins involved in cellular detoxification by catalyzing the conjugation of glutathione (GSH) with a wide range of endogenous and xenobiotic alkylating agents, including carcinogens, therapeutic drugs, environmental toxins and products of oxidative stress.
alr4046	alr4046	543	It consists of domains such as M61 glycyl aminopeptidase and domain belonging to peptidase Gluzincin family (thermolysin-like peptidases or TLPs)

Table 2: Physicohemical parameters of azoreductase accessory interactive proteins. The isoelectric point; molecular weight; negatively charged amino acids; positively charged amino acids; extinction coefficient; instability index; aliphatic index and GRAVY are described in separate columns for each of the accessory proteins.

Accessory interaction proteins	Isoelectric point	Molecular weight (MW)	-vely charged residues (-R)	+vely charged residues(+R)	Extinction coefficient	Instability Index(II)	Aliphatic Index (AI)	GRAVY
alr2599	7.86	14872.0	16	17	12615	59.00	92.98	-0.285
glyQ	5.09	34076.5	40	28	62590	42.52	88.23	-0.355
glyS	5.00	79346.3	94	73	62800	41.92	100.63	-0.176
alr4276	4.64	27154.3	39	25	21095	33.06	85.99	-0.282
alr2326	5.70	62589.0	62	51	110950	29.89	86.41	-0.329
ppc	5.46	117757.9	136	112	142140	50.65	93.30	-0.362
alr2106	9.10	12609.9	13	16	15470	43.94	103.64	-0.241
all1172	6.11	26514.7	30	25	26930	42.60	86.85	-0.439
alr1063	5.12	34880.6	40	24	55015	48.20	91.54	-0.205
alr4046	6.64	37783.0	37	37	67185	37.35	84.99	-0.307

Table 3: Global energy parameters of protein docking interaction of the accessory proteins with azo R. The different parameters such as attractive vanderwaal's (VdW), repulsive vanderwaal's (VdW), atomic contact energy (ATC) and hydrogen bonds (HB) are computed for each accessory protein. Alr2106 showing the best docking energy parameters followed by alr1063 and alr2326.

Azoreductase accessory interactive proteins	Global Energy	Attractive *VdW	Repulsive *VdW	*ACE	*HB
alr2599	-20.72	-30.16	22.62	4.85	-1.55
all4276	-34.14	-28.16	14.43	-0.76	-2.76
alr2326	-42.22	-29.62	15.67	-8.49	-5.59
alr2106	-61.42	-44.81	23.65	-5.01	-4.17
all1172	-26.89	-43.70	22.81	13.13	-3.15
alr1063	-32.96	-34.97	20.85	-7.12	-3.16
alr4046	-27.32	-32.53	9.53	2.86	-4.87