BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 9(12)

Computational protein structure modeling and analysis of UV-B stress protein in *Synechocystis* PCC 6803

Md Akhlaqur Rahman[1], Navaneet Chaturvedi[2], Sukrat Sinha[1], Paras Nath Pandey[3], Dwijendra Kumar Gupta[2], Shanthy Sundaram[1]* & Ashutosh Tripathi[1]

[1]Centre for Biotechnology, [2]Center of Bioinformatics, [3]Department of Mathematics, University of Allahabad, Allahabad-211002, India; Shanthy Sundaram – Email: shanthy_s@rediffmail.com; *Corresponding author

Received April 09, 2013; Accepted April 10, 2013; Published July 12, 2013

Abstract:
This study focuses on Ultra Violet stress (UVS) gene product which is a UV stress induced protein from cyanobacteria, *Synechocystis* PCC 6803. Three dimensional structural modeling of target UVS protein was carried out by homology modeling method. 3F2I pdb from *Nostoc* sp. PCC 7120 was selected as a suitable template protein structure. Ultimately, the detection of active binding regions was carried out for characterization of functional sites in modeled UV-B stress protein. The top five probable ligand binding sites were predicted and the common binding residues between target and template protein was analyzed. It has been validated for the first time that modeled UVS protein structure from *Synechocystis* PCC 6803 was structurally and functionally similar to well characterized UVS protein of another cyanobacterial species, *Nostoc sp* PCC 7120 because of having same structural motif and fold with similar protein topology and function. Investigations revealed that UVS protein from *Synechocystis* sp. might play significant role during ultraviolet resistance. Thus, it could be a potential biological source for remediation for UV induced stress.

Keywords: Cyanobacteria, Protein modeling, Hypothetical Protein, *Synechocystis* PCC 6803, Ultra Violet stress.

Background:
The continuous depletion of ozone layer results in an increased level of UVB radiation (280–315 nm) that reaches the Earth's surface and thereby increases the exposure of organisms in surface waters to UVB radiation **[1]**. The response of cyanobacteria to UVB radiation involves a series of damaging effects on DNA, protein and photosynthetic apparatus. It is well known that UVB radiation causes widespread damage in cells and also modulates the expression of many genes. Unlike transcriptomic approaches, only a few quantitative proteomic studies have explored the impact of damaging solar radiation on microorganisms. For example, a quantitative proteomic approach with 2D gels was used to compare the proteomes of irradiated and unirradiated *Deinococcus radiodurans* bacteria to identify the mechanisms of their extreme radio-resistance and DNA repair **[2, 3]** Protein breakdown and recycling, which depend on the levels of proteolytic enzymes, are an essential part of the plant response to environmental stress **[4]**. In response to environmental abiotic and biotic factors, cellular proteins are normally rebuilt. Degradation of damaged, misfolded and potentially harmful proteins provides free amino acids required for the synthesis of new proteins.

There are number of proteins related to UV-B stress which have not been reported. Identification, physiological and computational study of such differentially expressed proteins provides new ideas for future studies that will allow assessment of their physiological roles and significance in the acclimization

of cyanobacteria under UV-B stress condition. Some of them belong to the hypothetical and unknown group **[4]**. These hypothetical proteins are characterized by low identity to known, annotated proteins. Hypothetical proteins (HPs) lacking any significant sequence similarity to other ORFs in the databases are termed orphan ORFs (syn: ORFans) or ''poorly conserved ORFs''. About half the proteins in most genomes are candidates for HPs **[5]**. Among cyanobacteria, a lot of research shows that UV-B has a number of negative effects on cell physiology, damaging nucleic acids, protein and photosynthetic apparatus **[6].** So, it is essential to utilize proteomic and metabolic strategies to find a system level understanding of cyanobacterial response to UV-B stress. A number of new proteins are synthesized after prolonged UV-B radiation in cyanobacteia **[4]**. Some are known proteins while some of them are hypothetical proteins whose cellular function is not clear in relation to UV-B stress. An important example of this type of hypothetical protein is protease. A protease also termed as peptidase or proteinase, is an enzyme that conducts proteolysis. This proteolytic enzyme is found in all living organisms, and takes an important role in cell growth and differentiation. Protease is a valuable commercial enzyme and serves as a vital tool in determination of protein and polypeptides structure **[7]**. Detailed computational studies of protein sequence homology are essential for a variety of purposes and have, therefore, become routine in computational molecular biology and bioinformatics field. It is also seen that protein structure prediction is possible through bioinformatics. The functional analysis has been necessary to confirm such predictions.

The present study is an attempt to predict the structural aspects and putative binding sites of target hypothetical protein, protease from UV-B treated cyanobacteria, *Synechocystis* PCC 6803. This protein is identified by two- dimensional electrophoresis (2-DE) **[4]**.

Methodology:

Qing-yu Wu **[4]** had proposed a list of putative UV-B (short- and long- term) irradiation proteins from cyanobacterium, *Synechocystis* PCC 6803, identified by MALDI TOF/TOF. For our study, a UV-B stressed induced hypothetical protein, protease, was recruited from Gen Bank database **[8].** The target accessions no GI: 16331465 hypothetical proteins (protease) were selected from hypothetical category **[4].** The selection was made on the basis of finding suitable protein structure template from protein databank database **[9]**.

Three Dimensional Protein Modeling

To study the functional details of UV-B stress protein, it is required to have the three dimensional structure modeling of target protein of *Synechocystis* PCC 6803. BLASTp was applied to obtain a suitable template protein crystal structure. BLASTp **[10]** was run against protein databank **[9]** and the modeling was performed on SWISSMODEL **[11]**. After three dimensional modeling of target protein structure, it was necessary to validate the structure by the Ramachandran approach. Here, RAMPAGE **[12]** was used to validate the target protein structure.

The recognition of errors in experimental and theoretical models of protein structures is a major problem in structural biology. SWISSMODEL workspace provides the QMEAN4 score that is used to evaluate the generated target model protein. The global QMEAN4 scoring function **[13]** is a linear combination of four structural descriptors using statistical potentials. QMEAN4 is a reliability score for the whole model which can be used in order to compare and rank alternative models of the same target. The quality estimate ranges between 0 and 1 with higher values for better models. In addition, the QMEAN Z-score represents the measurement of the absolute quality of a model by providing an estimate of the 'degree of nativeness' of the structural features observed in a model and by describing the likelihood that a given model is of comparable quality to experimental structures. It was necessary to check the most similar protein structure from all entries of PDB; Dali server **[14]** was extensively used for this purpose. The Dali server is a network service for comparing proteins in 3D structure against the PDB.

Prediction of Binding Sites

Q-Site Finder was used to highlight the regions for ligand binding sites. It employs the interaction energy between the protein and a simple Van-der Waal probe to locate energetically favorable binding sites. Energetically favorable probe sites are clustered according to their spatial proximity and clusters are then ranked according to the sum of interaction energies for the sites within each cluster **[15]**. This method was widely applied for probable binding sites. Ligand binding sites were predicted for both target and template protein which depicted the common residues including common position between both proteins. This study indicates the residues playing a major role in cadmium binding sites.

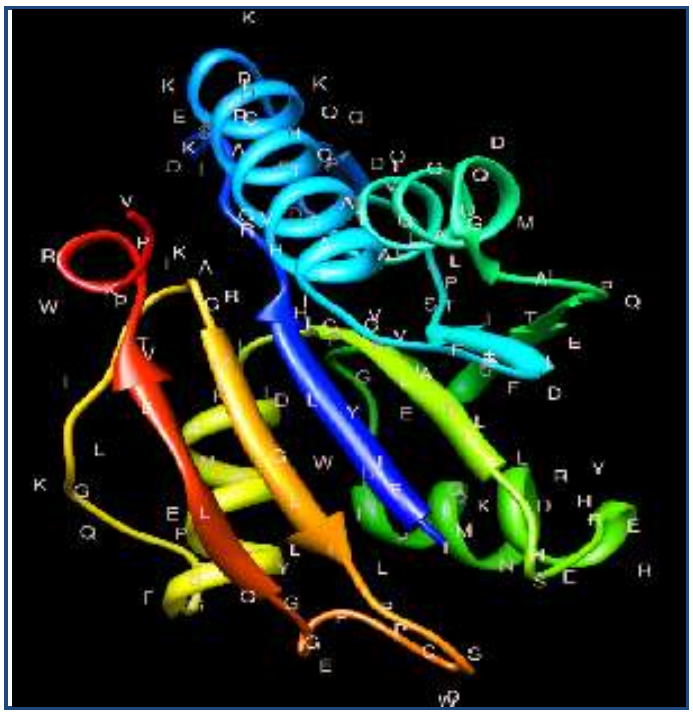

Figure 1: Three dimensional structural representation (Chimera view) ribbon display. Residues labeled by one-letter code.

Results & Discussion:

The target UV-B stress induced protein was obtained from NCBI protein database and the accession was noted as

NP_442193. BLASTp was applied against protein databank for cyanobacteria. The maximum identity (49%) for *Nostoc* sp. was observed among all obtained hits.

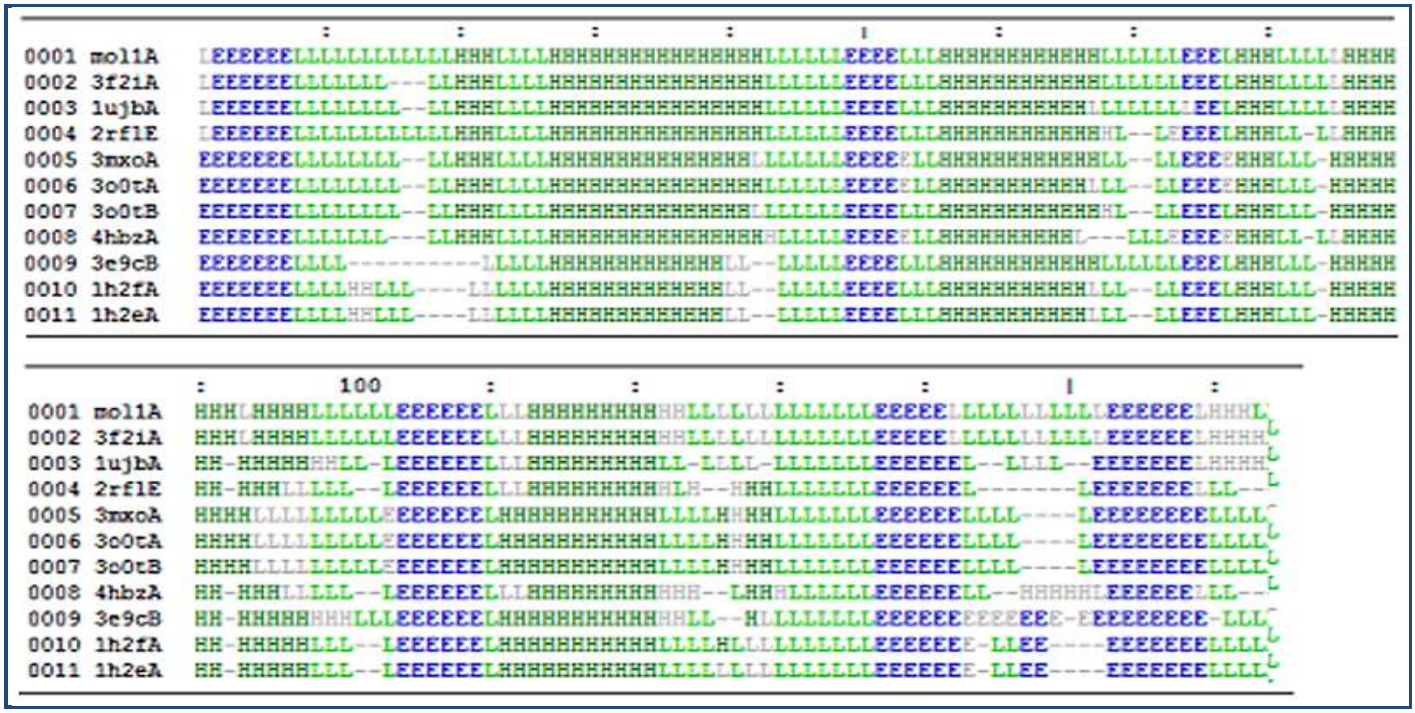

Figure 2: Alignment of sequences of target putative UV-B stress from *Synechocystis* PCC 6803, template from *Nostoc* sp. and other related protein sequences by DALI, showing consensus regions and secondary structure. Depiction of secondary structure assignments by DSSP (H/h: helix, E/e: strand, L/l: coil) in which he most frequent amino acid type is colored in each column. Mol1A is treated as target protein which is written at top row (0001 number).

Figure 3: Q-Site Finder analysis of target protein structure from *Synechocystis* PCC6803. The energetically most favored site with residues involved in interactions and estimated site volume; Total site volume: 15474.

Three Dimensional Protein Modeling

The target protein sequence of interest for *Synechocystis* PCC 6803 strain was obtained from NCBI protein database, accession, NP_442193. BLASTp produced significant alignments and 3F2I_A PDB, UV-B stress protein, from *Nostoc* sp, was chosen as a suitable template protein for computer modeling. All the inputs were studied using the SWISSMODEL for homology modeling. Three dimensional modeled protein structures are shown in **(Figure 1).** It consists of five alpha-beta domains and loop regions were embedded in between each of the alpha-beta domains.

Three dimensional structure of target protein of *Synechocystis* PCC 6803 strain was validated through RAMPAGE **(Figure 2)**. 156 residues (96.3%) were expected in favored region and only 6 (3.7%) numbers of residues were recruited in allowed region in Ramachandran contour plot. No residue was obtained in disallowed region. In context of QMEAN4 global scores, the Z-score indicates overall model quality and measures the deviation of the total energy of the modeled structure with respect to an energy distribution derived from random conformation. The overall QMEAN4 score was calculated as 0.63 and the Z-score was observed as -1.86. The QMEAN4 score is a composite score consisting of a linear combination of 4 statistical potential terms (estimated model reliability between 0-1) **Table 1 (see supplementary material)**. After complete refinement and assessment of predicted model, it has been found that the predicted model quality was good and reliable. After quantitative and qualitative analysis of predicted 3D structure model was successfully deposited to PMDB database **[16]** (PMDB id: PM0078697).

Similar Secondary Structure Analysis

Dali tool is extensively used to check the most similar protein structures from PDB. Three dimensional structures comparison is shown in **(Figure 3).** Here, the result describes the 3D structures that were obtained related to target UV-B protein structure. 10 PDB structures have been picked for comparing target protein. The Dali server results indicate that the structures having similarity above 70% were recruited for 3D-alignment. This study was needed to evaluate the secondary structure information of target protein structure. Dali result shows high quality secondary structure similarity from all 10 PDB crystal structures. Helix regions show better resemblance with target structure rather than sheets and coil regions. However, these regions also show better match with target protein **(Figure 2)**.

Prediction of Binding sites

Computational methods for the detection of active binding region and characterization of functional sites in protein structure have increasingly become an area of interest. There is at least one successful prediction in the top three predicted sites in 90% of proteins tested using Q-Site Finder. Generally, ligand binding site prediction method analyzes the protein surface for pockets. The ligand binding sites are usually in the largest interacting cavity having major active binding region. The cavities are defined by energetic criteria. The method calculates the Van-der Waal interaction energies of probe with protein. Q-Site finder depicts 10 major active binding cavities. Results are summarized in **(Figure 3)**. Most favorable binding sites contain amino acids with high conservation residue score.

Predicted binding site selection is highlighted according to the likelihood of being an actual binding site. The residues involved in each binding site for modeled protein were also tabulated in first three letters of name. In addition, the site volume of every predicted binding site was also taken into account. This site volume box gives information about the selected predicted binding site. It gives an estimate of cavity volume and total protein volume. All the predicted sites were selected which had the volume above 90 cubic angstroms.

We have chosen to analyze UVB protein from *Synechocystis* PCC 6803 by using tools and databases. The results indicate that all the amino acids are specific and there is an UVB stress protein domain in *Synechocystis* PCC 6803 that shows similarity with the *Nostoc* sp. In addition, comparative UVB stress protein modeling is done to construct a three dimensional model on one or more related proteins of known structure, which can be taken from the protein data bank **[9].** Further, for deducing the biological functions involved in the mechanism via structure function relationship, structure knowledge is essential for all areas of protein research such as enzyme kinetics, ligand–protein binding studies, gene characterization and construction, structure based molecule design, and rational designing of proteins **[17, 18]**. In addition, these models can speed up the process of experimental structure determination by molecular replacement phasing using X-ray crystallography. Active sites binding prediction was performed by Qsitefinder, which is widely used and is a reliable tool. The results indicate that the prediction can be applied in docking processes to understand the mechanism of protection against UV stress. Moreover, it could be helpful to enhance the tolerance to UV radiation stress.

Conclusion:

The present study will help to understand the mechanism of putative UV-B stress induced protein in cyanobacterium, *Synechocystis* PCC 6803 and the protective role of cyanobacteria in UV stress mechanisms. Identification of such proteases expressed under UV B stress will give clues related to the physiological activity and the importance of cyanobacteria, *Synechocystis* PCC 6803 acclimatization under UV stress. This piece of work has given an insight into the three dimensional protein structure, its putative binding site/s and their comparison with other putative UV stress induced proteins in other cyanobacteria.

References:

[1] Hader DP *et al. Photochem Photobiol Sci*. 2007 **6**: 267 [PMID: 17344962]
[2] Zhang C *et al. Proteomics.* 2005 **5**: 138 [PMID: 15593145]
[3] Basu B *et al. Mol Cell Proteomics*. 2012 **11**: M111.011734 [PMID: 21989019]
[4] Yang Gao *et al. J Exp Bot*. 2009 **60**: 1141 [PMID: 19261921]
[5] Minion FC *et al. J Bacteriol*. 2004 **186**: 7123 [PMID: 15489423]
[6] Caldwell MM *et al. Bioscience.* 1979 **29**: 520 DOI: 10.2307/1307719
[7] Rabbani Syed *et al. Bioinformation.* 2012 **8**: 175 [PMID: 22419836]

[8] Benson Da *et al. Nucleic Acids Res*. 2000 **28**:15 [PMID : 10592170]
[9] Bernstein FC *et al. J Mol Biol*. 1977 **112**: 535 [PMID: 875032]
[10] Altschul SF *et al. J Mol Biol*. 1990 **215**: 403 [PMID: 2231712]
[11] Arnold K *et al. Bioinformatics*. 2006 **112**: 195 [PMID: 16301204]
[12] Lovell SC *et al. Proteins*. 2002 **50**: 437 [PMID: 12557186]
[13] Benkert P *et al. Proteins*. 2008 **71**: 261 [PMID: 17932912]
[14] Holm L *et al. Nucleic Acids Res*. 2010 **38**: 545 [PMID: 20457744]
[15] Laurie AT *et al. Bioinformatics*. 2005 **21**: 190 [PMID: 15701681]
[16] Castrignanò T *et al. Nucleic Acids Res*. 2006 **34**: 306 [PMC: 16381873]
[17] Weibel ER *et al Proc Natl Acad Sci USA*. 1991 **88**: 10357 [PMID: 1946456]
[18] Hvidsten TR *et al. PLOS One*. 2009 **15**: 6266 [PMID: 19603073]

Edited by P Kangueane
Citation: **Rahman *et al.*** Bioinformation 9(12): 639-644 (2013)

BIOINFORMATION

open access

Supplementary material:

Table 1: The raw scores, Z-scores of the QMEAN composite score as well as all terms are provided relating to the quality estimates of the scores obtained for high-resolution reference structures solved experimentally by X-ray crystallography.

Scoring function term	Raw score	Z-score
C_beta interaction energy	-97.39	0.01
All-atom pairwise energy	-4359.64	-0.77
Solvation energy	-8.79	-0.99
Torsion angle energy	-27.61	-1.38
QMEAN4 score	0.634	-1.86