Isolation and computational characterization of glutathione peroxidase gene from an aquatic fern - Salvinia molesta

Sabdar Rahaman[1], Pankaj Kumar Singh[2], Protip Basu[2], Sohini Gupta[3],
Monojit Basu[4] and Sayak Ganguli[5]

[1]Department of Botany, Bangabasi Evening College, Kolkata
[2]Computational Biology Division, The Biome, Kolkata
[3]Post Graduate Department of Botany, Barasat Government College, Kolkata
[4]Block Development Office, Taldangra, District Bankura, West Bengal
[5]Amplicon Biosciences Pvt. Ltd. Palta - 743122,

Corresponding author: ganguli.sayak@gmail.com

Abstract. Pteridophytes and more specifically ferns represent a large but threatened group of plants which often serve as important environmental markers for pollution. Reports regarding stress responses in ferns are rare, apart from a few studies involving the ecological distribution and molecular marker studies. This work isolates a glutathione peroxidase enzyme from an aquatic fern widely distributed in fresh and polluted water bodies adjacent to sources of environmental polluted sources. Further computational analyses were performed to study the structure of the protein encoded by the open reading frame. Results indicate the presence of a large number of binding pockets which serve as important binding sites in the interactions with the cognate ligands.

Introduction:

Salvinia molesta D.S. Mitch.[1], is an aquatic fern commonly known as giant *Salvinia* or Kariba weed has been widely used for phytoremediation studies for cadmium and lead [2]. In this work we report the isolation and insilico characterization of glutathione peroxidase gene from an aquatic environment rich in lead. Theoretical modeling studies indicate that domain organization remains unchanged in environments with high lead content. Domains of the protein were isolated and subsequently modeled and simulated. Glutathione peroxidases belong to largest class of stress enzymes found across all domain of life. The unique feature of glutathione peroxidases is that they belong to category of selenoproteins. Earlier reports have enlightened the fact that plant glutathione peroxidases contain cysteine instead of selenocysteine which found across all domain of life [3].

Materials and Method:

DNA extraction, Amplification and Sequencing:

Plants were collected from different areas as described in Ganguli *et.al*[4] and DNA was extracted using the CTAB method[5,6] and genomic PCR was performed using genomic DNA. The gene was amplified using forward and reverse primers GCATAATCATATGACTGCCC and, AGAAAATTACAACAAATTCT respectively. Amplification was carried out for 30 cycles at standard operating conditions.

The capillary sequencing was done at the Indian Institute of Chemical Biology, Kolkata and the sequences were converted to FASTA format using Chromas 1.7 software[7]. The sequence was then submitted to GENBANK using the SEQUIN resource[8]. The sequence can be accessed with the unique accession number: EF620779.1.

Computational Analyses:

At first nucleotide sequence having accession no: EF620779.1 was retrieved from NCBI data warehouse. Homologues of this sequence were identified using Basic Local Alignment Search Tool[9] of NCBI which were further aligned and were subjected to Phylogenetic tree construction using Neighbor Joining algorithm[10]. The Phylogenetic tree was further analyzed. In the next step, corresponding Open Reading Frame for this sequence was generated using ORF Finder[11] tool. Next the 3D structure of the sequence was modeled using comparative modeling approach. The structure was further simulated using AMBER ff93 [12] and CHARMM force fields[13]. The fluctuation profiles of each of the residues were noted and graphically represented. The structures obtained were validated using PROCHECK[14] and QMEAN[15] and the model passing the respective threshold was selected for further analyses. Structure specific features such as cavities, Protrusions and Flats were calculated using amino acid specific Zernike's Descriptors with the help of server 3DSURFER[16]. Conformation specific binding sites calculations in the form of pocket identification was done based on background Voronoi diagrams which helped us to calculate the Delaunay Triangulation Scores leading to largest pocket identification[17]. The efficacy of ligand binding was examined by docking the generated structure with the two most common ligands of glutathione peroxidases - hydrogen peroxide and glutathione. Docking was performed using Autodock[18] and the ligand interactions were visualised using LIGPLOT[19].

Results

Distinct amplified product was obtained near the 500bp region and the subsequent sequencing revealed a 533 base pair long sequence.
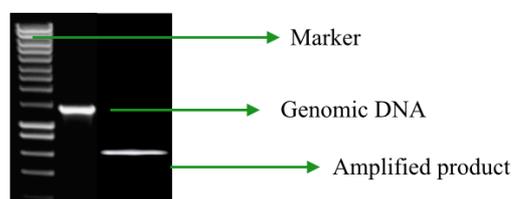

Figure 1: Products of Genomic DNA isolation and PCR amplification

This sequence was then subjected to the corresponding computational analyses pipeline as described above and distinct structural features were identified.
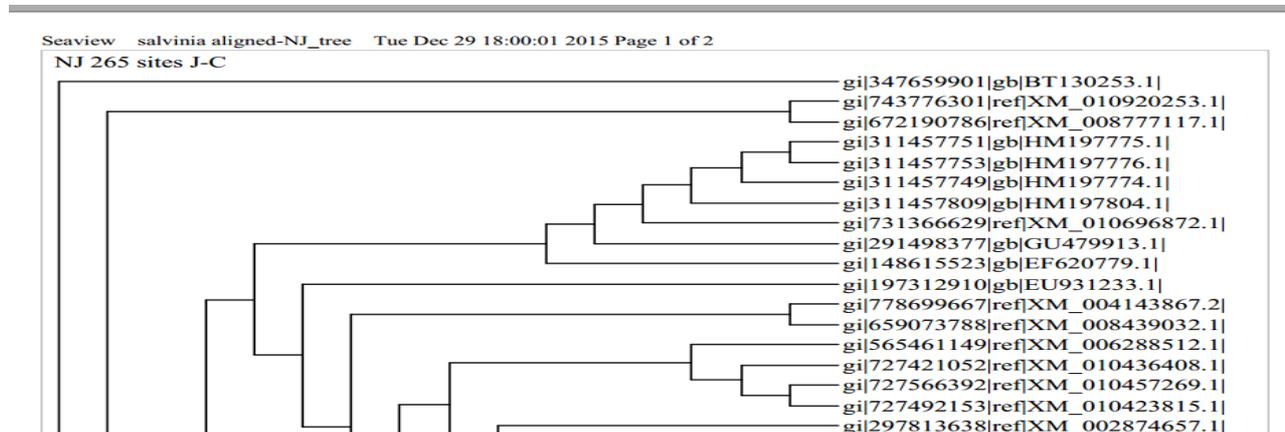

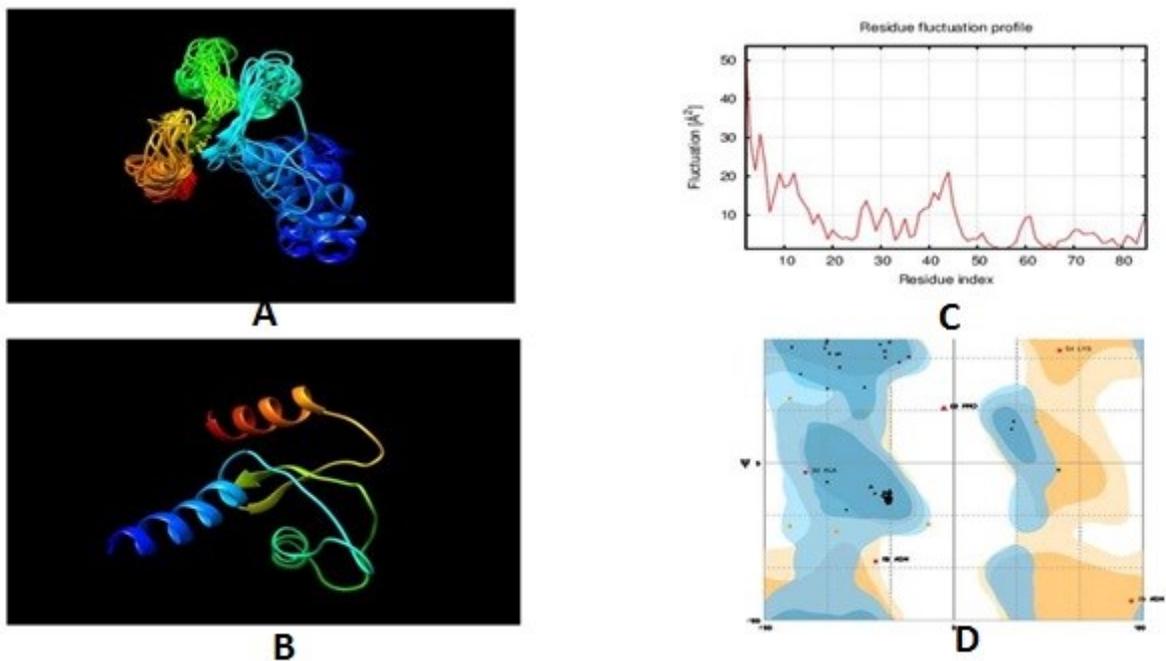**Figure 2:** Screenshot of Phylogenetic Tree of Homologues

Figure 3: Modeled protein (A) and its simulated structure (B) with respective Ramachandran plot (C) and residue fluctuation profiles (D).
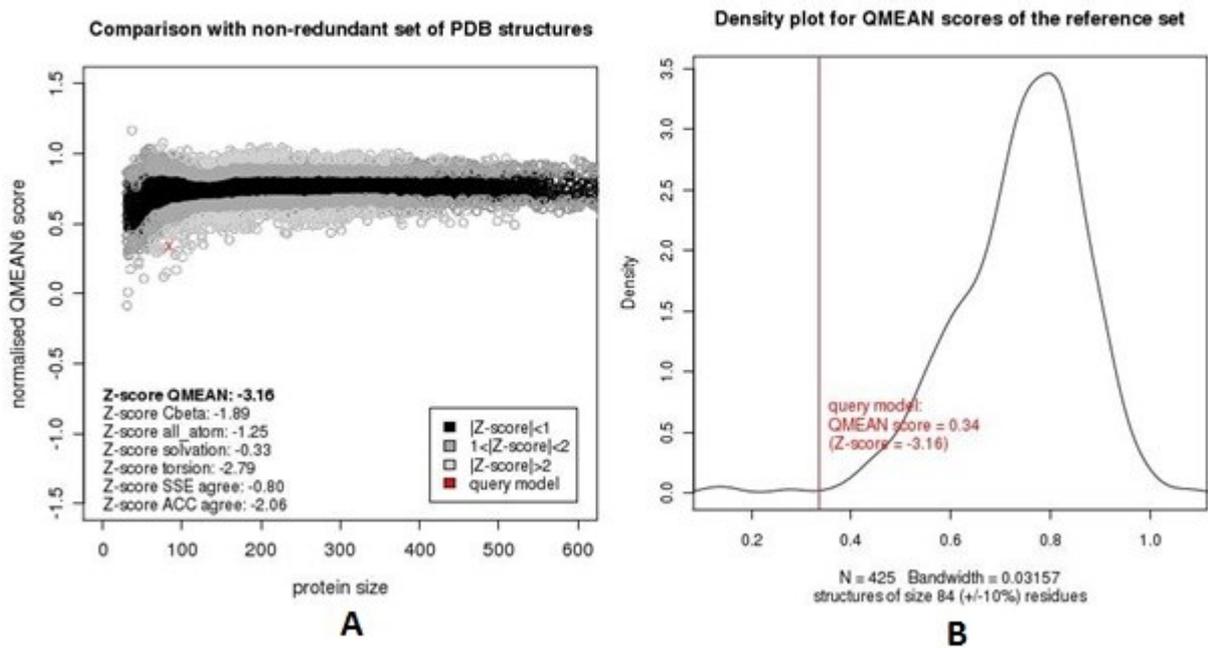

Figure 4: Validation of the simulated structure using QMEAN with corresponding Z score calculations (A) and Density (B).
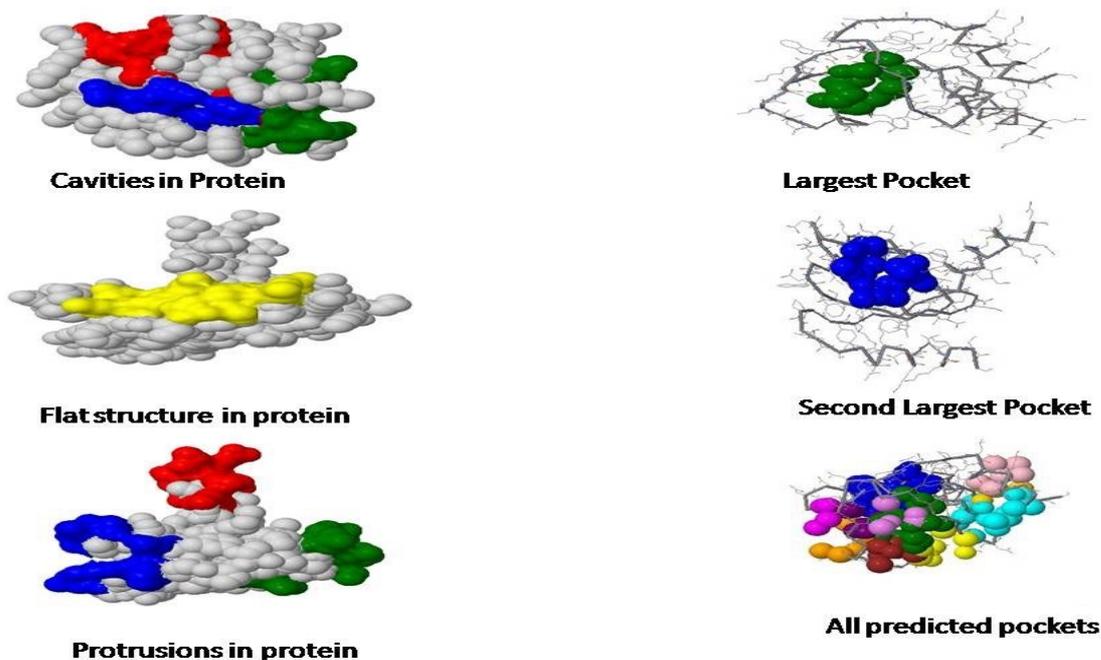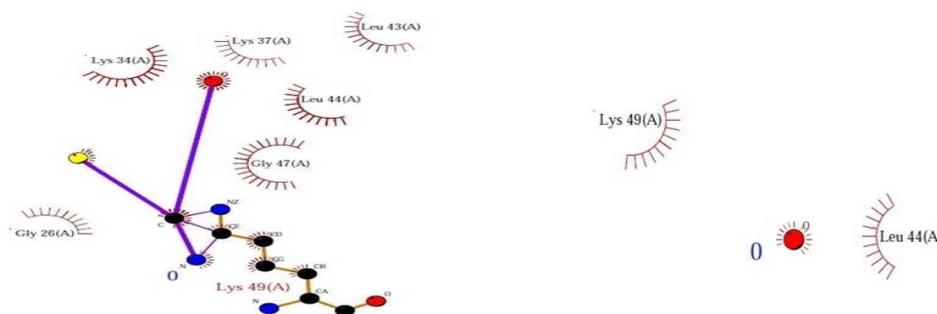
Figure 5: Identification of the feature based on protein structures - left hand column shows the cavities, protrusions and flats based on Zernike Descriptors; the right hand column shows the identification of pockets in the protein structures identified using Delaunay triangulation scores.

Figure 6: Protein - Ligand interactions represented graphically using Ligplot, left hand column displays interactions with glutathione while the right hand column shows interactions with hydrogen peroxide

Discussion:

The identification of the species was carried out according to the features suggested in and standard DNA isolation protocol was followed to identify the glutathione peroxidase gene using genomic PCR. The sequencing run revealed a 533 bp sequence of which the predicted coding region was of 267 nucleic acids. Thus this region was not a complete glutathione peroxidase enzyme but a portion of the entire product. Computational feature production identified the predicted coding region to be consisted of the glutathione peroxidase domain, a conserved feature for this family of enzymes. Comparative modelling and simulation revealed the presence of a robust domain topology and feature prediction using Zernike descriptors and Delaunay triangulation score revealed cavities and pockets in the predicted domain.

This domain was then used for the study of the interaction between the two most established ligands of this family of enzymes, - glutathione and hydrogen peroxide. Lysine 49 was identified to be the common interaction site for both the ligands which is a testimony to the observations made by Ganguli *et.al.*[20], regarding the conserved structural features and their participation in ligand interactions in this enzyme family.

Conclusion

This work reports for the first time the predicted structure of a stress associated enzyme from *Salvinia molesta* and its subsequent interactions with cognate ligands.

Acknowledgement:

References:

[1]   D.S. Mitchell, Brit. Fern Gaz. (1972) 10(5): 251-252.

[2]   S. Phetsombat,M. Kruatrachue, P. Pokethitiyook,S. Upatham , J Environ Biol. (2006)(4):645-52.

[3]   S. Ganguli,S. Gupta, A. Bhowmick, J. Ansari and M. Basu, Indian Science Cruiser (2008) page 38-45

[4]   S. Ganguli, A. Datta, International Letters of Natural Sciences, Vol. 7,(2013),pp. 49-62,

[5]   J.A. Bryant, DNA extraction, p. 1-12.In P.M. Dey and J.B. Harborne (Eds.), Methods in Plant Biochemistry, (1997),Vol. 10b. Academic Press, San Diego

[6]   C.S. Kim,C.H. Lee, J.S. Shin, Y.S. Chungand, N.I. Hyung, NucleicAcids Res. (1997) 25:1085-1086.

[7]   P. Technelysium, Chromas lite 2.01. (2007), http://www. technelysium.com.au/chromas_lite. html

[8]   M.D. Bethesda, National Center for Biotechnology Information (US); (2011)-. Available from: http://www.ncbi.nlm.nih.gov/books/NBK51157/

[9]   F. S. Altschul, , W. Gish,W. Miller,W.E. Myers, and J.D. Lipman, (1990) J. Mol. Biol. 215:403-410

[10]   N. Saitou , M. Nei ,Molecular Biology and Evolution, 4(4), (1987) pp. 406-425,

[11]   ORF Finder, http://www.ncbi.nlm.nih.gov/gorf/gorf.html

[12]   W.D. Cornell,P. Cieplak,C.I. Bayly,I.R. Gould,K.M. Merz, D.M. Ferguson,D.C. Spellmeyer,T Fox, J.W. Caldwell,P.A. Kollman,J. Am. Chem. Soc. (1995)**117**: 5179–5197.

[13]   B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus.

[14]   J. Comp. Chem. 4, (1983),187-217

[15]   R.A. Laskowski, M.W. MacArthur, D.S. Moss ,J.M. Thornton, J. App. Cryst., **26**, (1993), 283-291

[16]   P. Benkert ,M. Künzli,T. Schwede , Nucleic Acids Res. 1;37 (2009):W510-4

[17]   D. La, J. E. Rodriguez, V. Venkatraman, B. Li, L. Sael, S. Ueng, S. Ahrendt, and D. Kihara. Bioinformatics 25: (2009)2843-2844.

[18]   J. Dundas, Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz, and J. Liang., Nucleic Acid Research, 34: (2006),W116-W118.

[19]   G. M.Morris, R.Huey,W. Lindstrom, M. F. Sanner,R.K. Belew,D.S. Goodsell and A.J. Olson, J. Computational Chemistry (2009), 16: 2785-91

[20]   A.C. Wallace, R .A.Laskowski, J.M. Thornton, Protein Eng., 8, (1996). 127-134.

[21]   S.Ganguli,A.Datta, American Journal of Bioinformatics Research (2015), 5(1): 9-15