ISSN 0976 – 3333

Available Online at www.ijpba.info

International Journal of Pharmaceutical & Biological Archives 2013; 4(4): 753 - 757

ORIGINAL RESEARCH ARTICLE

In Silico Sequence Analysis, Homology Modelling and Functional Annotation of Toxin I Hypothetical Protein of Catfish, *Plotosus lineatus*

T. R. Barathkumar[1], Jeganathan Manivannan[2], Muthusamy Thangaraj*[1]

[1]*Centre of Advanced Study in Marine Biology, Faculty of Marine Sciences, Annamalai University, Parangipetai – 608 502, Tamil Nadu, India*

[2]*Department of Biochemistry, Annamalai University, Annamalai Nagar – 608 002, Tamil Nadu, India*

Received 11 May 2013; Revised 14 August 2013; Accepted 25 Aug 2013

ABSTRACT

Plotosus lineatus, also known as oriental catfish has been investigated for the characterization of proteinaceous toxins secreted from the venom glands in dorsal and pectoral spines. Particularly, a toxin (Toxin I) was characterized and reported to have more potential lethal, edema-forming and nociceptive activities and its primary structure was only elucidated. This toxin was highly homologous with other fish natterin-like proteins. Since, the unavailability of the 3D structure of natterin-like protein, this attempt was carried out and Homology modelling was performed using Phyre3D-PSSM folding server. In the present study, subcellular localization prediction suggested that it is a cytoplasmic protein. Quality analysis of the model indicated that it is a reliable model. Furthermore, it was discovered that this proteinaceous toxin (Toxin I) is involved in two biological processes, pathogenesis and inter-species interaction between the organisms and the biochemical function of the protein is hydrolase activity, acting on the glycosyl bonds.

Key words: *Plotosus lineatus*, natterin-like protein, Homology modeling, Functional annotation.

1. INTRODUCTION

A number of fishes are venomous and have pungent spines with venom glands, usually in fins. Catfishes, members of the order Siluriformes, are representative venomous fish having venom glands in dorsal and pectoral spines that are locked into place when threatened. As many as 1250 – 1625 species of catfishes have been estimated to be venomous [1]. On envenomation by catfishes, various local symptoms such as intense pain, edema and erythema are induced in victims; systemic symptoms such as tachycardia, weakness and vomiting are frequently observed and even fatalities have been reported in severe cases [2]. Extracts from the catfish spines have been established to be toxic [3-5]. However, the catfish venom gland toxins are extremely unstable proteins as in other venomous fish and hence have not been purified, except for the Indian catfish *Plotosus canius* toxin (toxin-PC) of 15 kDa that is lethal and cardiotoxic, having neuromuscular blocking activity [4]. It should be noted that some venomous catfishes, such as Arabian Gulf catfish *Arius bilineatus* (formerly identified as *Arius thalassinus*) [6], Madamango sea catfish *Cathorops spixii* [5] and oriental catfish *Plotosus lineatus* [7,8], contain proteinaceous toxins in the skin secretion as well as in the venom gland. This leads us to assume that the local and systemic symptoms induced by some catfishes are the combined effects of venom gland toxin and skin toxin. In this respect, for a better understanding of the symptoms upon envenomation by catfishes, it is important to clarify detailed properties and structure of the skin toxin.

The information of 3D structure remains an indispensable fact for experimentally discovering the functionality of any protein. This is partly due to the considerable experimental challenge and manual inputs required to solve three dimensional structures by methods such as X-ray diffraction and multi-dimensional nuclear magnetic resonance (NMR) spectroscopy in comparison to high-throughput sequencing [9]. Moreover, the rate at which protein sequence data is accumulating is far more than the structural information available,

***Corresponding Author:** Muthusamy Thangaraj, **Email:** coralholder@yahoo.com

thus creating a gap between available sequences and experimentally solved structures. Computational methods like homology modeling can help reduce this gap. It is known that existing proteins are result of continuous evolution of previously existing ones, thus proteins can be grouped into families [10]. Homology modeling methods use the fact that evolutionary related proteins share a similar structure. Therefore, models of a protein with unknown structure (target) can bebuilt based on an alignment of a protein of known structure (template). This typically involves four steps [11,12]: (1) identification of homologs that can be used as template(s) for modeling; (2) alignment of the target sequence to the template(s); (3) building a model for the target based on the information from the alignment(s); and (4) evaluation of the model. Finally, all four steps can be repeated until a satisfactory model is obtained.

2. MATERIALS AND METHODS

Protein retrieval and sequence analysis:

The cDNA sequence of protein was retrieved from NCBI genbank database using accession no. BAK19070. which is then translated to amino acid by similarity search in blastx. Physiochemical properties of the predicted protein were computed by ProtParam tool (web.expasy.org/protparam/). The parameters computed by ProtParam included the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY). Sub-cellular localization of any protein is important understanding protein function. Prediction of sub-cellular localization of protein was carried out by CELLO v.2.5[13,14].

Computational modelling

Predict Protein [15] was employed for computing and analyzing the secondary structural features of Toxin-I amino acid sequence. A three-dimensional model of test protein was generated by homology modelling [16] (Phyre; 3D-PSSM folding server ;). In brief, this method aligns a test sequence to one or more template structures with known structures as determined by crystallization/X-ray diffraction, or NMR spectrometry. This method allows for the identification of homology based on PSI-BLAST alignments in combination with a profile–profile matching algorithm [15] which adjusts for secondary structure alignments. Along with this hypothetical protein, non-self identifying (i.e., non-sequence identify) comparator templates with greatest homology to the target toxin I toxin sequence were haemolytic lectin (*Laetiporus sulphureus*; 45% sequence coverage; PDB code: 1w3g), mannose/glucose-specific lectin (44% sequence coverage; PDB code: 1zgs), and myrosinase binding protein (*Arabidopsis thaliana*; 45% sequence coverage; PDB code: 2jz4). These structures each had 100% confidence and served as controls for homology modeling of toxin I protein.

Quality and reliability assessments

Once the 3D model was generated, energy minimization was performed by GROMOS96 force field in a Swiss-PdbViewer. Structural evaluation and stereochemical analyses were performed using ProSA-web[17,18] displaying Z-scores and Procheck [19] visualising Ramachandran plot. Furthermore, superimposition of query and template structure, and visualization of generated models was performed using UCSF Chimera 1.5.3.

Function annotations of the protein

To functionally annotate the toxin I predicted protein, Profunc was used, and to find the conserved domains in protein to identify its family, it was searched against close orthologous family members. NCBI Conserved Domain Database (NCBI CDD) [20] was used to find the conserved domains or ancient domains in the protein sequence.

3. RESULTS AND DISCUSSION

The present study was to perform sequence and structure analysis of *Plotosus lineatus* hypothetical toxin I protein. The protein sequence was translated and cDNA sequence retrieved using accession no. BAK19070 from Genbank database.

Protein sequence analysis

ProtParam was used to find out the physiochemical properties from protein sequence. The hypothetical protein was predicted to have 317 amino acids, with molecular weight of 34737 Daltons and theoretical isoelectric point (PI) of 6.25. An isoelectric point below 7 indicates a negatively charged protein. The instability index (II) is computed to be 23.04. This classifies the protein as stable. The N-terminal of the sequence considered is M (Methionine). Therefore estimated half-life is 30 hours (mammalian reticulocytes, *in-vitro*), >20 hours (yeast, *in-vivo*)

and >10 hours (Escherichia coli, *in-vivo*). The negative Grand average of hydropathicity (GRAVY) of value -0.292 indicates that the protein is hydrophilic and soluble in nature. Threonine and glycine were found in rich amounts in the protein.

Cellular functions are often localized in specific compartments; therefore, predicting the subcellular localization of unknown proteins can give information about their functions and can also help in understanding disease mechanisms and developing drugs. The subcellular localization prediction using CELLO predicted that our protein is a cytoplasmic protein and this protein does not contain a nuclear localization signal. PredictProtein was used to predict the secondary structure of the protein. Results showed that protein is a mixed protein having composition of Strand = 64.04%, Loop = 35.96%.Ramachandran plot and Errat analysis revealed that the tested hypothetical protein consists rich amount of amino acids which posess significance for the predicted protein would be highly flexible (**Fig 4 & 5**).

3D structure prediction using homology modelling approach

Protein 3D structure is very important in understanding the protein interactions, functions and their localization. Homology modeling is the most common structure prediction method. The multidimensional lectin signature consists of sequence and structural elements distributed in a hallmark pattern in 3-dimensional space. At the structural level, the lectin protein is typically characterized by the presence of a strand including loops. To assess whether hypothetical protein contained these structural components, molecular modeling analyses were carried out to generate a predictive model of toxin I protein secondary structure. For this analysis, hemolytic lectin was identified as the template for 3-dimensional modeling. Several physicochemical features of the predicted 3-dimensional structure of toxin I were homologous to those in glycoproteins. The molecule appears to consist of two distinct regions, and an extended region that is relatively unstructured (**Fig 1**). From these perspectives, the predicted overall fold of toxin-I is consistent with that of haemolytic lectin. ProSA uses knowledge-based potentials of CA atoms. The Z-score of -3.66 indicates the overall model quality of Toxin-I (**Fig 2**). Z-score also measures the deviation of total energy of the structure with respect to an energy distribution derived from random conformations. The scores indicate a highly reliable structure and are well within the range of scores typically found for proteins of similar size. The energy plot shows the local model quality by plotting knowledge-based energies as a function of amino acid sequence position (**Fig 3**).

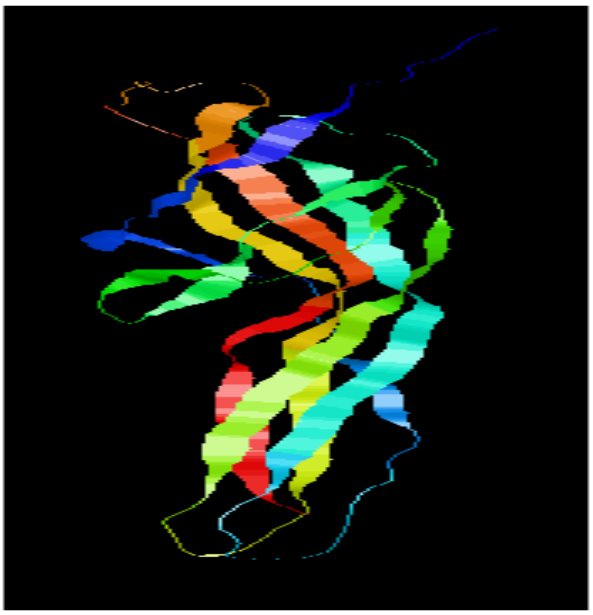

Fig 1: Predicted 3D structure of *P. lineatus* hypothetical protein BAK19070

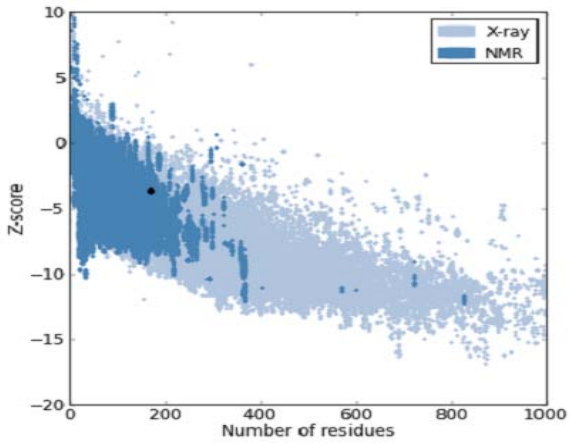

Fig 2: Z-score of natterin-like protein using PROSA web

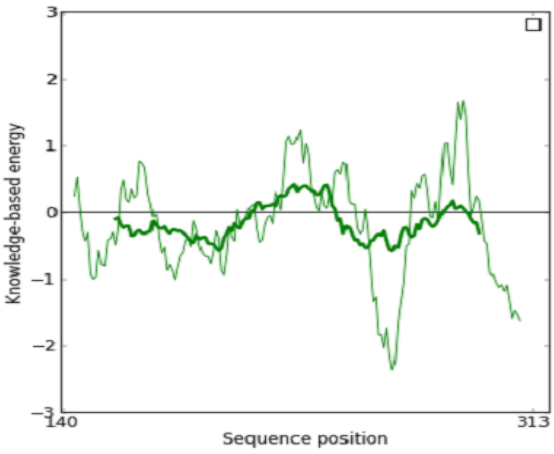

Fig. 3 Knowledge-based energy of natterin-like protein using PROSA web

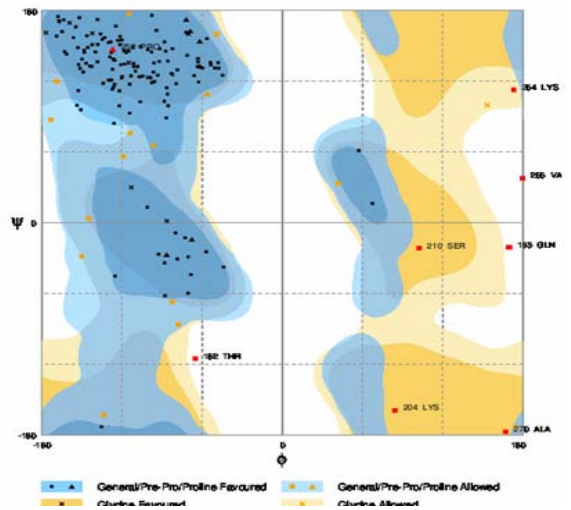

Fig 4: Ramachandran plot assessment for defined regions of flexibility by Rampage analysis

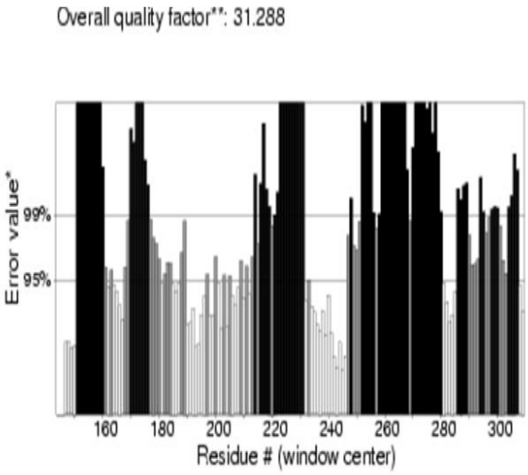

Fig. 5 Overall quality factor checked by ERRAT

Table 1: Amino acid frequency and composition of test protein

Amino acid	Frequency	Percentage (%)
Ala(A)	16	5
Arg(R)	9	2.8
Asn(N)	16	5
Asp(D)	11	3.5
Cys(C)	4	1.3
Gln(Q)	6	1.9
Glu(E)	23	7.3
Gly(G)	32	10.1
His(H)	5	1.6
Ile(I)	18	5.7
Leu(L)	16	5
Lys(K)	23	7.3
Met(M)	6	1.9
Phe(F)	18	5.7
Pro(P)	9	2.8
Ser(S)	29	9.1
Thr(T)	36	11.4
Trp(W)	7	2.2
Tyr(Y)	9	2.8
Val(V)	24	7.6

Functional annotation of the protein

To hypothetically annotate the function of the *P. lineatus* hypothetical protein BAK19070 ProFunc was used. It was discovered that protein is involved in two biological processes, pathogenesis and inter-species interaction between the organisms and the biochemical function of the protein is hydrolase activity, acting on the glycosyl bonds. To further investigate about the function of protein by finding its family; it was searched in the NCBI Conserved Domain Database (NCBI CDD) to find conserved domains so that its family can be identified.

The results showed that *P. lineatus* hypothetical protein BAK19070 has Jacalin-like lectin domain and belongs to Jacalin-like super family. Jacalin-like lectins are sugar-binding protein domains mostly found in plants. They adopt a beta-prism topology consistent with a circularly permuted three-fold repeat of a structural motif. Proteins containing this domain may bind mono- or oligosaccharides with high specificity. The domain can exist in tandem-repeat arrangements with up to six copies, and in architectures combined with a variety of other functional domains. Taxonomic distribution is not restricted to plants, the domain is also found in various mammalian proteins, for example.

5. CONCLUSION

Our main objective of this study was to perform sequence analysis, structure analysis and homology modeling on *P. lineatus* hypothetical protein BAK19070. We have used various sequence and structure analysis tools that helped in understanding of the sequence and its structure. Furthermore, protein was functionally annotated by using ProFunc and by searching conserved domain of the protein. As a part of present study, we used homology modeling approach to propose the first 3D structure of the *P. lineatus* hypothetical protein. The predicted 3D structure will provide more insight in understanding the structure and function of the protein. Moreover, this structure can be used for drug designing or understanding the interactions between proteins.

Declaration of interest

The authors report no declarations of interest

REFERENCE

1. Wright, J.J., 2009. Diversity, phylogenetic distribution, and origins of venomous catfishes. BMC Evol. Biol. 9, 282–293.

2. Halstead, B.W., 1970. Poisonous and Venomous Marine Animals of the World, vol. 3. U. S. Government Printing Office, Washington, D. C.
3. Birkhead, W.S., 1972. Toxicity of stings of ariid and ictalurid catfishes. Copeia 1972, 790–807.
4. Auddy, B., Muhuri, D.C., Alam, M.I., Gomes, A., 1995. A lethal protein toxin (toxin-PC) from the Indian catfish (Plotosus canius, Hamilton) venom. Nat. Toxins 3, 363–368.
5. Junqueira, M.E., Grund, L.Z., Orii, N.M., Saraiva, T.C., de Magalhães Lopes, C.A., Lima, C., Lopes-Ferreira, M., 2007. Analysis of the inflammatory reaction induced by the catfish (Cathorops spixii) venoms. Toxicon 49, 909–919.
6. Thomson, M., Al-Hassan, J.M., Fayad, S., Al-Saleh, J., Ali, M., 1998. Purification of a toxic factor from Arabian Gulf catfish epidermal secretions. Toxicon 36, 859–866.
7. Shiomi, K., Takamiya, M., Yamanaka, H., Kikuchi, T., Konno, K., 1986. Hemolytic, lethal and edema-forming activities of the skin secretion from the oriental catfish (Plotosus lineatus). Toxicon 24, 1015–1018.
8. Shiomi, K., Takamiya, M., Yamanaka, H., Kikuchi, T., 1987. Purification of a lethal factor in the skin secretion from the oriental catfish Plotosus lineatus. Nippon Suisan Gakkaishi 53, 1275–1280.
9. Sheehan D, Siobhán O'Sullivan (2011). Online homology modelling as a means of bridging the sequence-structure gap. Bioengineered. 2(6):299-305.
10. Giancarlo R., A. Siragusa, E. Siragusa, F. Utro (2007). A Basic Analysis Toolkit for Biological Sequences, Algorithms for Molecular Biology, 2(10), 404-406.
11. Sánchez, R. and Sali, A. 1997. Advances in comparative protein-structure modeling. Curr. Opin. Struct. Biol. 7: 206–214
12. Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sánchez, R., Melo, F., and Sali, A.2000. Comparative protein structure modeling of genes and genomes. Annu.Rev. Biophys. Biomol. Struct. 29: 291–325
13. Yu C. S., C. J. Lin, J. K. Hwang (2006). Predicting Subcellular Localization of Proteins for Gram-negative Bacteria by Support Vector Machines based on n-peptide Compositions, Protein Science, 13, 1402-1406.
14. Yu C. S., Y. C. Chen, C. H. Lu, J. K. Hwang (2004). Prediction of Protein Subcellular Localization, Proteins: Structure, Function and Bioinformatics, 64, 643-651.
15. Rost B., G. Yachdav, J. Liu (2004). The PredictProtein Server, Nucleic Acids Research, 32(Web Server issue), W321-W326.
16. Kelley LA, MacCallum RM, Sternberg MJ (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. J. Mol. Biol., 299, pp. 499–520
17. Wiederstein & Sippl (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research 35, W407-W410.
18. Sippl, M.J. (1993). Recognition of Errors in Three-Dimensional structures of proteins. Proteins 17, 355-362.
19. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992). Stereochemical quality of protein structure coordinates. Proteins, 12, 345-364
20. Marchler-Bauer A., S. H. Bryant (2004). CD-Search: Protein Domain Annotations on the Fly, Nucleic Acids Research, 32(W), 327-331.