Available online at www.ijpab.com

DOI: http://dx.doi.org/10.18782/2320-7051.6354

ISSN: 2320 – 7051

Int. J. Pure App. Biosci. **6 (1):** 1392-1398 (2018)

Research Article

International Journal of Pure & Applied Bioscience

OPEN ACCESS

In-Silco Analysis of Fish Antifreeze Proteins and their Physicochemical Characterization

Mukesh Kumar[1*], Hanzel Saldana[2], Roshan Kumar Ram[1], Himanshu Bhattacharyya[1], Norine D' Souza[2]

[1]Fish Genetics and Biotechnology Division, ICAR- Central Institute of Fisheries Education, Versova, Mumbai, Maharashtra 400061

[2]Department of Biotechnology, St. Xavier College, Mahapalika Marg, Mumbai, Maharashtra 400001

*Corresponding Author E-mail: mukesh.fbtpa503@cife.edu.in

Received: 5.01.2018 | Revised: 13.02.2018 | Accepted: 18.02.2018

ABSTRACT

Antifreeze protein's (AFP) are proteins which have an important function of ice binding and growth inhibition of ice crystals and found in plants, bacteria, fungi, invertebrates and fish. In this present study 17 Sequence of fish antifreeze proteins were selected for their Physiochemical properties, Secondary structure, Homology modeling, validation, and for functional analysis. Insilco tools were used to describe the physiochemical, functional and structural properties of these protein. Physico-chemical properties like as molecular weight, amino acid composition, isoelectric point, charge residue, Grand average of hydropathicity, instability index are computed .The studies give insight into fish antifreeze protein structure.

Key words: *Antifreeze protein, Fish, Isoelectric point, Ice binding, Physiochemical properties.*

INTRODUCTION

Fishes can survive under seawater below their freezing point of their blood because of antifreeze proteins. Antifreeze proteins have helped many researchers to identify the various uses of AFP as cryo-protectants in preservation of various biological samples[1]. Fish antifreeze proteins consist of 2 categories- antifreeze glycoproteins and antifreeze proteins. It constitutes 3.4% of the blood of the total marine fishes. They are classified as Type I, II, III and IV on the basis of their discovery. Type I (3.3-4.5Kda) is simple in structure and identified by alanine rich residue (alpha helix) comprising more than 60% and also its various repeating sequence which starts with threonine. It can be differentiated from the others by the presence of its carbohydrate recognition domain which are very much similar to calcium dependant lectins[2]. Type II consist of globular proteins that having high amount of disulphide bonds[3]. Type III also has globular proteins which has a beta sandwich structure having 8 beta strands. Its size is intermediate as compared to other types and has no differentiating features that differs them from the remaining types of AFP[4].

Cite this article: Kumar, M., Saldana, H., Ram, R.K., Bhattacharyya, H. and D.' Souza, N., In-Silco Analysis of Fish Antifreeze Proteins and Their Physicochemical Characterization, *Int. J. Pure App. Biosci.* **6(1):** 1392-1398 (2018). doi: http://dx.doi.org/10.18782/2320-7051.6354

Type IV is identified on the basis of its high glutamine content which is 17% and it has an important feature that it consist of 4 alpha helix bundles that consist of 60% of the total proteins[5]. The main aim in studying antifreeze proteins is its wide application in various fields of biology such as cryo surgery and various tissue explanations[6]. It has various applications in pyrexia, hypothermia etc[7]. In agriculture biology, tolerance of the crops can be increased with the use of plant AFP[6]. In food biology it has a major role like enhancing the durability of yoghurt, various chilled meat products etc[8]. Computational methods for analysis of the protein help us to understand various properties of the protein and its structural analysis. In this study, we will focus on the in silco characterization fish antifreeze proteins along with its physiochemical and structural properties and also see its applications.

MATERIAL AND METHODS

A total of 17 Sequence of antifreeze proteins were selected and protein fasta file was retrieved using NCBI and UNIPROT database and were used for further analysis. (Table 1)

Physiochemical properties:

Physiochemical properties were calculated using Protparam tool[9] which gives details about molecular weight, theoretical Pi, amino acid composition etc. The amino acid composition was determined by using protparam tool. (Table 2).

Functional analysis and secondary structure analysis:

Transmembrane regions were identified using SOSUI server[10] (Table 3). SOPMA[11] and GOR[12] were used for predicting secondary structure of the protein (table 4). In these tools the presence of alpha helx, pi helix, Beta Bridge, beta turn etc. is determined in terms of percentage. Analysing ss-bonding of cysteine residues and identifying various disulphide bridges in the particular protein was done by CYS_REC[13]. The prediction of the protein structure was done using ITASSER[14] which characterises various structural based templates from PDB database by various approaches that have various atomic based models. Protein ligand binding site were identified using COACH server[15] (Table 5). Ramchandran plot for the respective protein structure were determined by RAMPAGE[16] and the quality of the Ramchandran plot was characterised by WHAT IF server17 (Table 6). Coil server[18] is a program that is used to compare a sequence that is consisting of parallel two-stranded coiled-coils and helps to determine a similarity score.

Homology modeling and validation:

The homology modelling was carried out using- ITASSER[14] and Swiss model[19] (https://swissmodel.expasy.org/) which is a programmed homology modelling server used for visualising protein structures and saving them in PDB format and using it by other tools for analysis and CHIMERA[20] (https://www.cgl.ucsf.edu/chimera/) helps in visualisation and analysis of various structures. It also has certain actions through which the structures can be edited and saved in PDB format and can be used for various analysis.

Functional properties:

METAPOCKET2.0 (http://projects.biotec.tudresden.de/metapocket/index.php) were characterised by determining the distance between ligand binding site and all the protein residues[21]. MOTIFF FINDER (http://www.genome.jp/tools/motif/) was done in order to determine at which position of your protein is the motif present and description of the motif is given[22]. PSORT (https://psort.hgc.jp/) determines various protein signals and various sites in the sequence provided[23]. NETOGLYC (http://www.cbs.dtu.dk/services/NetOGlyc/) helps in creating a branch of predictions of mucin gal NAC-O-GLYCOSYLATION site the respective proteins[24]. SIGNAL P (http://www.cbs.dtu.dk/services/SignalP/) identifies the direction of cleavage sites from different protein sequences[25]. PROSITE (http://prosite.expasy.org/) consists of various entries consisting of various protein domains, families and functional sites and various patterns and profiles which helps to identify them[26].

Predict protein (https://www.predictprotein.org/) helps to predict various structural and functional annotations. (Included in supplementary table)

RESULTS AND DISCUSSION

Anti-Freeze protein (AFPs) have been isolated and characterized from different organisms likes as fish, fungi, bacteria, insects, and plants etc. These are structurally diverse group of proteins and different in their structures, from primary structure to tertiary structure, and they bind to ice crystals in order to reduce the freezing point of water[27]. These are ice-binding proteins that prevent ice growing through the depression of the freezing point of a solution to below the melting point. This difference between freezing and melting point is called as thermal hysteresis and occur due to adsorption of the AFP on the crystal surface of ice. This interaction results ice growth to take place in a convex surface between adjacent AFPs, thus decreasing the freezing point[28]. In this present Study 17 Antifreeze protein sequences of fishes were retrieved from NCBI and UNIPROT database and computationally analysed. The primary structure analysis which means the Physico-chemical properties of AFPs from these fish varieties are computed using Expasy's ProtParam tool and shown in Table-2. By analysing these AFPs we found that high non-polar residues are abundant in these proteins due to this it seems that these AFPs be mostly hydrophobic ,this result is very similar with study by Hossain[29]. The isolelectric point (pI) is the value at which the molecule carries no charges or the and positive negative charges is found to be equal. At a pH below their pI value of proteins carry a net positive charge; above their pI these carry a net negative charge. In our study this is highest for (11.75- Liparis gibbus) and lowest value for (4.41- Limanda ferruginea) respectively. The Extinction coefficient (EC) at 280nm wavelength ranging from 1490 to 41480 at 280nm wavelength. The high extinction coeficient value in the antifreeze proteins interprets that there is high concentration of Cys, Trp and Tyr. The calculated extinction coeficient values help in various quantitative studies of protein-protein and protein-ligand interactions in solution. The instability index value of antifreeze proteins was computed by EsPasy protparam which gives an approximation of the stability of the protein in vitro. A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable. The instability indexes of AFPs are ranging from 10.56 to 47.99. The aliphatic index (AI) which is the relative volume of a protein occupied by aliphatic side chains (A, V, I and L) is regarded as a positive factor for the increase of thermal stability of globular proteins. The lower thermal stability values indicats a more flexible structure when we compare with other AFPs. The very high aliphatic index infers that these AFPs may be stable for a wide range of temperature. The Grand Average Hydropathy (GRAVY) is a phenomenon used for calculating the hydropathy value of all the amino acids upon the number of residues in the sequences. GRAVY concludes that almost all fish antifreeze proteins analysed in this study is hydrophobic (Table 2). Low Grand Average Hydropathy index of antifreeze proteins shows the possibility of better interaction with water. We also performed including trans-membrane (TM) region identification, prediction of disulphide bonding pairs etc. The SOSUI server helps in the identification of various transmembrane helices with their corresponding length and it helps to distinguish between membrane proteins from stable proteins. The server SOSUI classifies AAA49465.1, AHZ08737.1, J7I8W2, P09031, NP_001038953.1 as membrane proteins and the rest as soluble proteins. The secondary structures of AFPs were predicted by SOPMA (Self optimized prediction method with alignment) and GOR. This secondary structure indicates whether a given amino acid lies in a helix, strand or coil. CYS_REC inteprets the presence of S-S bonds and their possible bonding pairs among all Cys residues. Possible disulphide bond pairing and patterns with probability were predicted by CYS_REC from

primary sequence and S-S bonds were identified from 3D structure. Ramchandarn plot validation of is done and allowed and disallowed region is calculated. All value which are not reflected inresult table is attached with supplementary table. Prosite helps to find various sites like N-myristoylation site, Amidation site and cAMP- and cGMP-dependent protein kinase phosphorylation site etc. Predict protein gives 3 types of biological pathway. It explains how the protein has application in cellular, molecular and biological process. For study the receptor-ligand interaction the identification of ligand binding sites is an important parameter and there are no individual methods that can provide the optimal prediction for all proteins. In this present study we use TM-SITE and S-SITE for protein–ligand binding site predictions. We get Siganl P value positive for 9 protein out of seventeen protein taken in this study.

Table 1: Antifreeze proteins retrieved from NCBI and UNIPROT database

ACCESSION NUMBER	SEQUENCE DESCRIPTION	ORGANISM
A0A0G1Z3E6	Ice-structuring protein	Scleropages formosus
A0A088AZS1	Ice-structuring protein	Lycodichthys dearborni
A0A088AZS4	Ice-structuring protein	Lycodichthys dearborni
AHZ08737.1	type IV antifreeze protein	Carassius gibelio
NP_001038953.1	antifreeze protein type IV precursor	Danio rerio
J7I2Q7	Ice-structuring protein	Pholis gunnellus
J7I8W2	Ice-structuring protein	Ulvaria subbifurcata
AAR22529.1	type I antifreeze protein	Liparis gibbus
ADU02183.1	type IV antifreeze protein	Notothenia coriiceps
A0A060VSB7	Ice-structuring protein	Oncorhynchus mykiss
AAA49442.1	antifreeze protein precursor	Osmerus mordax
P09031	Ice-structuring protein	Limanda ferruginea
ADU02182.1	type-IV antifreeze protein	Pleuragramma antarctica
A0A0S7LB01	Ice-structuring protein	Poeciliopsis prolifica
AEI59129.1	antifreeze protein	Tautogolabrus adspersus
AAA49465.1	antifreeze protein precursor	Pseudopleuronectes americanus
ALL26680.1	type 3 antifreeze protein	Zoarces americanus

Table 2: Physicochemical properties of AFPs from different Fish varieties are computed using Expasy's ProtParam tool

Accession number	Length	Molecular Weight	PI Isoelectric point	(-) R Negative charged Residue	(+) R Positive charged Residue	Extinction coefficient	Instability index	Aliphatic index	c
AHZ08737.1	129	14300.58	5.57	16	13	2980	46.42	107.36	0.012
NP_001038953.1	130	14194.45	5.15	16	13	2980	47.99	101.37	-0.032
AAR22529.1	113	9741.96	11.75	1	8	-	15.26	81.59	0.736
ADU02183.1	128	14118.17	4.7	15	10	1490	30.66	104.61	-0.102
AAA49442.1	175	19053.95	5.16	16	9	41480	33.3	76.46	0.171
ADU02182.1	128	142025.2	4.72	15	9	1490	39.74	100.78	-0.128
AEI59129.1	47	4240.8	9.3	4	6	-	16.66	69.79	0.16
AAA49465.1	91	8326.45	5.92	4	4	5500	13.86	88.02	0.826
ALL26680.1	74	7941.47	8.99	4	6	1490	30.16	89.46	0.172
A0A088AZS1	291	31758.88	5.4	29	26	10680	22.19	104.78	0.218
A0A088AZS4	227	24679.48	5.9	21	20	7575	21.19	108.55	0.278
A0A060VSB7	268	29990.51	7.01	36	36	19160	36.67	73.47	-0.381
A0A0S7LB01	214	23363.1	6.91	28	28	13200	40.03	94.21	-0.145
J7I2Q7	90	9618.55	9.03	5	7	1490	19.2	108.33	0.492
J7I8W2	90	9442.19	6.04	5	5	1490	30.62	121.22	0.693
A0A0G1Z3E6	70	7349.54	4.92	11	9	3105	12.84	108.43	0.127
P09031	97	8864.95	4.41	8	5	5500	10.56	89.69	0.739

Table 3 Transmembrane protein results

Accession no	Transmembrane region (N terminal –C terminal)	Type	Length
AAA49465.1	MALSLFTVGQLIFLFWTLRIT	Primary	21
	AAKAAPAAVADPAAAAAAAVADT	Secondary	23
AHZ08737.1	MKFFLIAVLVTALAIGSESVSLV	Primary	23
J7I8W2	MNSVIFTGLVFVLLCVDNMSSAA	Primary	23
	LIPINTALTLVMMRAEVVSPLGI	Secondary	23
P09031	AATATAAAAAAASAAAAAAATTA	Secondary	23
	AKVSAGAAATAAAAVVAAKNAAT	Primary	23
NP_001038953.1	FSLIAVIVVALAIGSESASLV	Primary	21

Table 4: Secondary structure prediction by SOPMA and GOR (GOR results in bold)

Accession Number	Alpha helix		Extended		Beta turn		Random coil	
A0A0G1Z3E6	52.86	**28.57**	18.57	**20.00**	15.71	**0.0**	12.86	**51.43**
AFPIII-13	52.58	**34.36**	8.93	**17.53**	5.15	**0.0**	33.33	**48.11**
AFPIII-24	32.16	**32.16**	20.26	**20.26**	0.0	**0.0**	47.58	**47.58**
AHZ08737.1	82.17	**72.09**	4.65	**6.20**	4.65	**0.0**	8.53	**21.71**
NP_001038953.1	82.31	**74.62**	5.38	**10.00**	3.08	**0.0**	9.23	**15.38**
J7I2Q7	42.22	**22.22**	12.22	**23.33**	2.22	**0.0**	43.33	**54.44**
J7I8W2	45.56	**33.33**	12.22	**13.33**	10.00	**0.0**	32.22	**53.33**
AAR22529.1	86.73	**84.07**	2.65	**4.42**	0.0	**0.0**	10.62	**11.5**
ADU02183.1	82.81	**87.50**	2.34	**0.0**	0.0	**1.56**	14.84	**10.94**
A0A060VSB7	36.94	**35.07**	19.03	**22.01**	0.0	**10.45**	44.03	**32.46**
AAA49442.1	22.29	**30.29**	25.14	**21.71**	0.0	**8.57**	52.57	**39.43**
P09031	80.41	**80.41**	3.09	**3.09**	0.0	**0.0**	16.49	**16.49**
ADU02182.1	82.81	**82.81**	2.34	**2.34**	0.0	**0.0**	14.84	**14.84**
A0A0S7LB01	35.98	**35.98**	15.89	**15.89**	0.0	**0.0**	48.13	**48.13**
AEI59129.1	82.98	**82.98**	0.0	**0.0**	0.0	**0.0**	17.02	**17.02**
AAA49465.1	61.54	**61.54**	16.48	**10.48**	0.0	**0.0**	21.98	**21.98**
ALL26680.1	29.73	**29.73**	20.27	**20.27**	0.0	**0.0**	50.00	**50.0**

Table 5: Coach results (Based on the rank 1 scores) for A0A0G1Z3E6

Method	C score	Prediction binding site
TM-SITE	0.23	39,42
S-SITE	0.20	5,6,7,8,29,30,32,33,34,35
COFACTOR	0.01	19,23
FIND-SITE	0.47	41,42,58,65,66,67,68
CONCAVITY	0.07	5,6,7,26,28,29,30,31,32,33,53,55,56,57

Table 6: What IF Results

ACCESSION NO	Z-SCORE (I tasser)	Z-SCORE (COACH)
A0A0G1Z3E6	-1.654	-1.161
AFPIII-13	-2.831	-4.158
AHZ08737.1	-3.215	-5.098
NP_001038953.1	-1.355	-1.030
J7I2Q7	1.544	0.360
J7I8W2	-1.902	-2.106
ADU02183.1	-3.084	-1.249
A0A060VSB7	2.096	4.158
AAA49442.1	-0.703	3.055
P09031	-2.553	-0.701
ADU02182.1	4.805	-1.151
A0A0S7LB01	0.507	4.543
AEI59129.1	-0.486	-0.111
AAA49465.1	3.972	-0.764
ALL26680.1	-1.651	2.085
A0A0G1Z3E6	-1.654	1.660
AFPIII-13	-2.831	-2.975

Acknowledgement

The authors acknowledge the Director, ICAR-CIFE, for providing the necessary facilities for the conduct of this research work.

REFERENCES

1. Gilbert, J.A., Davies, P.L., Laybourn-Parry J. A hyperactive, Ca2+-dependent antifreeze protein in an Antarctic bacterium. FEMS Microbiology Letters, **245(1):** 67-72 (2005).
2. Gwak, I.G., sic Jung, W., Kim, H.J., Kang, S.H., Jin, E., Antifreeze protein in Antarctic marine diatom, Chaetoceros neogracile. Marine biotechnology, **12(6):** 630-9 (2010).
3. Hanada, Y., Nishimiya, Y., Miura, A., Tsuda, S., Kondo, H., Hyperactive antifreeze protein from an Antarctic sea ice bacterium Colwellia sp. has a compound ice-binding site without repetitive sequences. The FEBS journal, **281(16):** 3576-90 (2014).
4. Bayer-Giraldi, M., Weikusat I., Besir, H., Dieckmann, G., Characterization of an antifreeze protein from the polar diatom Fragilariopsis cylindrus and its relevance in sea ice. Cryobiology, **63(3):** 210-9 (2011).
5. Uhlig, C., Kabisch, J., Palm, G.J., Valentin, K., Schweder, T., Krell, A., Heterologous expression, refolding and functional characterization of two antifreeze proteins from Fragilariopsis cylindrus (Bacillariophyceae). Cryobiology, **63(3):** 220-8 (2011).
6. Chao, H., Houston, M.E., Hodges, R.S., Kay, C.M., Sykes, B.D, Loewen, M.C., Davies PL, Sönnichsen F.D. A diminished role for hydrogen bonds in antifreeze protein binding to ice. Biochemistry, **36(48):** 14652-60 (1997).
7. Koushafar, H., Pham, L., Lee, C., Rubinsky, B., Chemical adjuvant cryosurgery with antifreeze proteins. *Journal of surgical oncology*, **66(2):** 114-21 (1997).
8. Griffith, M., Ewart, K.V., Antifreeze proteins and their potential use in frozen foods. Biotechnology advances, **13(3):** 375-402 (1995).
9. Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R, Appel, R.D., Bairoch, A. Protein identification and analysis tools on the ExPASy server. In the proteomics protocols handbook 571-607 (2005).
10. Hirokawa, T, Boon-Chieng, S., Mitaku, S., SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics (Oxford, England). 1998 **14(4):** 378-9 (1998).
11. Geourjon, C. and Deleage, G., SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Bioinformatics, **11(6):** 681-684 (1995).
12. Heymann, M., Paramelle, D., Subra, G., Forest, E., Martinez, J., Geourjon, C., Deléage, G., MSX-3D: a tool to validate 3D protein models using mass spectrometry. Bioinformatics, **24(23):** 2782-3(2008).
13. Roy, S., Maheshwari, N., Chauhan, R., Sen, N.K., Sharma, A., Structure prediction and functional characterization of secondary metabolite proteins of Ocimum. Bioinformation, **6(8):** 315 (2011).
14. Zhang, C., Mortuza, S.M., He, B., Wang, Y., Zhang, Y., Template-based and free modeling of I-TASSER and QUARK pipelines using predicted contact maps in CASP12. Proteins: Structure, Function, and Bioinformatics. (2017).
15. He, B., Mortuza, S.M., Wang, Y., Shen, H.B, Zhang, Y., NeBcon: protein contact map prediction using neural network training coupled with naïve Bayes classifiers. Bioinformatics, **33(15):** 2296-306 (2017).
16. Batut, P., Gingeras, T.R., RAMPAGE: Promoter Activity Profiling by Paired-End Sequencing of 5′-Complete cDNAs. Current protocols in molecular biology, **11:25** (2013).
17. Vriend, G., WHAT IF: a molecular modeling and drug design program.

Journal of molecular graphics, **8(1):** 52-61 (1990).

18. Lupas, A., Van Dyke, M., and Stock, J. Predicting Coiled Coils from Protein Sequences, Science, **252:** 1162-1164 (1991).
19. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., Schwede, T., SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic acids research. **42:** 252-8 (2014).
20. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, **25(13):**1605-12 (2004).
21. Zhang, Z., Li, Y., Lin, B., Schroeder, M., Huang, B., Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. Bioinformatics, **27(15):** 2083-8 (2011).
22. Bailey, T.L, Boden, M., Buske, F.A, Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S., MEME SUITE: tools for motif discovery and searching. Nucleic acids research, **37:** 202-8 (2009).
23. Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., Nakai, K., WoLF PSORT: protein localization predictor. Nucleic acids research, **35:** 585-7 (2007).
24. Steentoft, C., Vakhrushev, S.Y., Joshi, H.J., Kong, Y., Vester-Christensen, M.B., Katrine, T., Schjoldager, B.G., Lavrsen, K., Dabelsteen, S., Pedersen, N.B., Marcos-Silva, L., Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *The EMBO journal*, **32(10):** 1478-88 (2013).
25. Nielsen, H., Predicting secretory proteins with SignalP. Protein Function Prediction: Methods and Protocols, 59-73 (2017).
26. Sigrist, C.J.A., de Castro, E., Cerutti, L., Cuche, B.A., Hulo, N., Bridge, A., Bougueleret, L., Xenarios, I., New and continuing developments at PROSITE Nucleic Acids Res. (2012).
27. Davies, P.L., Baardsnes, J., Kuiper, M.J., Walker, V.K., Structure and function of antifreeze proteins. Philos Trans R Soc Lond B. **357:** 927–935 (2002).
28. Kristiansen, E., Zachariassen, K.E., The mechanism by which fish antifreeze proteins cause thermal hysteresis. Cryobiology. **51:** 262–280 (2005).
29. Hossain, M.M., Fish antifreeze proteins: Computational analysis and physicochemical characterization. *Inte. Current Pharmaceutical Journal*, **1(2):** 18-26 (2012).