Hossain, International Current Pharmaceutical Journal 2012, 1(2): 18-26
http://www.icpjonline.com/documents/Vol1Issue2/01.pdf

International Current Pharmaceutical Journal

ORIGINAL RESEARCH ARTICLE OPEN ACCESS

Fish antifreeze proteins: Computational analysis and physicochemical characterization

*Md. Musharaf Hossain

Department of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh

ABSTRACT

Antifreeze proteins (AFPs) protect organisms from freezing and shows great diversity in structure, and they have been found in a variety of organisms. In this study, a total of 15 antifreeze proteins of fish were selected where they represent distinct physicochemical and structural features. The present paper uses bioinformatics approach to describe the physiochemical, functional and structural properties of Antifreeze proteins. Several Physico-chemical properties such as pI, EC, AI, GRAVY and instability index are computed and provide data about these proteins and their properties. The result of primary structure analysis infers that, fish antifreeze proteins are mostly hydrophobic. Disulfide bridges and secondary structures were analyzed using CYS_REC and SOPMA respectively. The three dimensional structure of Antifreeze proteins is predicted by using three homology modeling server Geno3D, Swiss-model and CPHmodels. The model was evaluated with PROCHECK, WHAT IF, and ProSA programs. Model visualization and analysis was done with Pymol. These structures will provide a good foundation for functional analysis of experimentally derived crystal structures.

***Key Words*:** Antifreeze Proteins, Computational tools, hydrophobicity, homology modeling, isoelectric point.

INTRODUCTION

Many organisms living in cold environments can survive subzero temperature by providing antifreeze protein or antifreeze glycoprotein, where they inhibit the growth of ice by possessing thermal hysteresis (TH) or ice crystallization inhibition (RI) activity. AFPs protect the organisms from freezing at temperature below 1°C by binding with ice crystals and modify their growth through an adsorption-inhibition mechanism (Raymond *et al*., 1977). Through this unique technique, they protect themselves from cell membrane damage and some other harmful physical and chemical changes. Though, AFPs were first identified in fishes (Fletcher *et al*., 2001), they also have been found in plants (Griffith *et al*., 2004), fungi (Hoshino *et al*., 2003) and bacterial species (Kawahara *et al*., 2004; Gilbert *et al*., 2004; Gilbert *et al*., 2005). Beside their diversified sources various structurally distinct AFPs have evolved independently (Davies and Sykes, 1997). A total of 5 structurally distinct antifreeze proteins are identified in fish so far and classified as Antifreeze glycoprotein (AFGP) and antifreeze protein type I, type II, type III, and type IV based on their distinct physicochemical and structural features (Davies *et al*., 1990). Antifreeze activity of AFPs attracts a lot of attention due to their wide potential commercial applications including preservation, transgenic production (Wang *et al*., 1995) and cryosurgery. AFPs have potential applications in agriculture for the production of economically valuable fishes against low temperature. Other proposed applications of AFPs are found in cryosurgery of tumors, transplantation, transfusion (Fletcher *et al*., 1999) and as a component of ice-cream to prevent the formation of hard and large ice crystals (FSANZ, 2006). Many researchers are working for many years on antifreeze protein and they have purified and analyzed AFPs from different sources to resolve the protein-ice interaction (Madura *et al*., 2000; Jorov *et al*., 2004), evolution of AFPs (Lui *et al*., 2007; Sandve *et al*., 2008; Deng *et al*., 2010), structure function co-rrelation (Graether *et al*., 2004), molecular dynamics

***Corresponding Author**:
Md. Musharaf Hossain
Department of Genetic Engineering and Biotechnology
Shahjalal University of Science and Technology
Sylhet-3114, Bangladesh
E-mail: *musharaf27@gmail.com*
Contact No.: +880 1814 800 818

© 2012 Hossain; licensee Saki Publishing Club. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nd/3.0/), which permits unrestricted use (including commercial use), distribution and reproduction of the work in any medium, provided the original work is properly cited and remain unaltered.

Table 1: Antifreeze protein sequences retrieved from Swiss-Prot database.

Accession no	Sequence description	Organism
P11920	Ice-structuring glycoprotein	*Eleginus gracilis* (Saffron cod)
Q01758	Type-2 ice-structuring protein	*Osmerus mordax* (Rainbow smelt)
Q92006	Type III Antifreeze protein	*Rhigophila dearborni* (Antarctic eelpout)
Q1AMQ4	Type III antifreeze protein	*Pachycara brachycephalum* (Antarctic eelpout)
Q1AMR1	Type II antifreeze protein	*Clupea harengus* (Atlantic herring)
Q1AMQ2	Type III antifreeze protein	*Anarhichas minor* (Arctic spotted wolffish)
P84493	Type II antifreeze protein	*Hypomesus nipponensis* (Japanese smelt)
Q1AMQ8	Type III antifreeze protein	*Macrozoarces americanus* (Ocean pout)
B1P0S1	Type I hyperactive AFP	*Pseudopleuronectes americanus* (Winter flounder)
Q1AMQ1	Type IV antifreeze protein	*M. octodecimspinosis* (Longhorn sculpin)
Q1AMQ6	Type III antifreeze protein	*Rhigophila dearborni* (Antarctic eelpout)
Q1AMR0	Type II antifreeze protein	*Osmerus mordax* (Rainbow smelt)
Q1AMR3	Type I antifreeze protein	*Pseudopleuronectes americanus* (Winter flounder)
Q1AMQ9	Type III antifreeze protein	*Macrozoarces americanus* (Ocean pout)
Q1AMQ3	Type III antifreeze protein	*Anarhichas minor* (Arctic spotted wolffish)

and modeling studies (Lin *et al*., 2007). Besides all aspects of experimental analysis, now-a-days several computational approaches and online servers provide great opportunities for the characterization and analysis of protein to accelerate experimental approaches as well as widening scientific thoughts. Computational tools provide researchers a cost effective way to understand physicochemical and the structural properties of a protein for the successful design of many biological experiments with in a short range of time. Several physicochemical properties of a protein such as molecular weight, grand average hydropathy (GRAVY), aliphatic index (AI), extinction coefficient (EC), isolelectric point (p^I), instability index (II) etc. can be computed along with their functional characterization. Numerous structure and function studies of AFPs have been reported experimentally from time to time while computational study of AFPs are much more limited. So, the effort has been taken to study the physicochemical and structural properties of AFPs from fishes. In this study, we will focus on the in silico characterization and homology modeling of AFPs from different fish varieties.

MATERIALS AND METHODS

Sequences of Antifreeze protein were retrieved from Swiss-Prot, a public domain protein database (Boeckmann *et al*., 2003). A total of 15 sequences of fish were retrieved from Swiss-Prot by random selection. Table 1 shows the protein sequences considered in this study. All antifreeze protein sequences were retrieved in FASTA format and used for further analysis.

Physicochemical properties

The physicochemical properties were calculated from the primary structure of antifreeze protein where the physicochemical parameters, theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient (Gill and Von Hippel, 1989), half-life (Tobias *et al*., 1991), instability index (Guruprasad *et al*., 1990), aliphatic index (Ikai *et al*., 1980) and grand average hydrophathy (GRAVY) (Kyte and Doolottle, 1982) were computed using the Expasy's ProtParam (Gasteiger *et al*., 2005) (http://us.expasy.org /tools/protparam.html) prediction server. The amino acid compositions of all retrieved protein sequences were also determined (Table 2) and the physicochemical properties were tabulated in table 3.

Functional characterization and secondary structure analysis

The identification of transmembrane regions of a protein was identified by SOSUI server. Table 4 represents the transmembrane regions identified for those antifreeze proteins. The predicted transmembrane helices were visualized and analyzed using Helical wheel Plots. SOPMA (Geourjon and

Table 2: Amino acid composition of fish antifreeze proteins (in percentage) computed using EsPasy tool.

AMINO ACIDS	Q1AMR0	Q1AMR3	Q1AMQ9	Q1AMQ3	P11920	Q01758	Q92006	Q1AMQ4	Q1AMR1	Q1AMQ2	P84493	Q1AMQ8	B1P0S1	Q1AMQ1	Q1AMQ6
Ala	10.9%	45.1%	5.7%	10.2%	57.9%	10.9%	8.0%	8.1%	8.8%	10.2%	8.8%	5.7%	56.4%	11.7%	8.0%
Arg	1.1%	2.4%	2.3%	3.4%	5.3%	1.1%	2.3%	1.2%	2.0%	2.3%	1.4%	1.1%	0.5%	1.6%	1.1%
Asn	4.0%	2.4%	3.4%	4.5%	0.0%	4.0%	5.7%	3.5%	2.7%	4.5%	2.0%	4.5%	2.3%	3.9%	5.7%
Asp	5.1%	3.7%	3.4%	3.4%	0.0%	5.1%	3.4%	3.5%	6.1%	3.4%	6.8%	1.1%	3.2%	4.7%	3.4%
Cys	6.3%	0.0%	1.1%	1.1%	0.0%	6.3%	1.1%	1.2%	7.5%	1.1%	7.5%	1.1%	0.0%	0.0%	1.1%
Gln	2.3%	1.2%	4.5%	4.5%	0.0%	2.3%	2.3%	3.5%	3.4%	4.5%	3.4%	4.5%	0.5%	14.8%	2.3%
Glu	4.0%	2.4%	2.3%	1.1%	0.0%	4.0%	3.4%	4.7%	4.8%	1.1%	4.1%	3.4%	1.4%	7.8%	3.4%
Gly	6.3%	2.4%	4.5%	5.7%	0.0%	6.3%	4.6%	3.5%	4.8%	5.7%	5.4%	6.8%	0.5%	2.3%	4.6%
His	3.4%	0.0%	1.1%	2.3%	0.0%	3.4%	1.1%	1.2%	2.7%	2.3%	2.7%	1.1%	0.0%	0.8%	1.1%
Ile	4.6%	2.4%	6.8%	10.2%	0.0%	4.6%	8.0%	4.7%	5.4%	10.2%	2.0%	8.0%	5.0%	7.8%	5.7%
Leu	8.0%	7.3%	13.6%	12.5%	0.0%	8.0%	11.5%	10.5%	9.5%	11.4%	9.5%	9.1%	2.3%	10.2%	12.6%
Lys	4.0%	3.7%	2.3%	2.3%	0.0%	4.0%	5.7%	9.3%	4.1%	3.4%	3.4%	6.8%	3.7%	7.0%	6.9%
Met	4.6%	2.4%	8.0%	8.0%	0.0%	4.6%	9.2%	9.3%	4.1%	6.8%	4.1%	9.1%	0.9%	4.7%	10.3%
Phe	3.4%	3.7%	1.1%	1.1%	0.0%	4.0%	1.1%	1.2%	3.4%	1.1%	3.4%	2.3%	2.3%	3.9%	1.1%
Pro	5.1%	6.1%	8.0%	8.0%	10.5%	5.1%	6.9%	5.8%	4.1%	9.1%	3.4%	8.0%	0.9%	2.3%	6.9%
Ser	6.9%	3.7%	5.7%	4.5%	0.0%	6.3%	4.6%	8.1%	8.8%	4.5%	8.2%	5.7%	5.5%	3.1%	4.6%
Thr	9.1%	8.5%	11.4%	6.8%	26.3%	9.1%	6.9%	7.0%	8.2%	8.0%	11.0%	8.0%	9.6%	7.0%	6.9%
Trp	4.0%	1.2%	0.0%	0.0%	0.0%	4.0%	0.0%	0.0%	4.8%	0.0%	4.8%	0.0%	0.5%	0.0%	0.0%
Tyr	1.1%	0.0%	2.3%	1.1%	0.0%	1.1%	1.1%	1.2%	1.4%	1.1%	2.0%	1.1%	0.5%	0.8%	1.1%
Val	5.7%	1.2%	12.5%	9.1%	0.0%	5.7%	12.6%	12.8%	3.4%	9.1%	5.4%	12.5%	4.1%	5.5%	12.6%

Deleage, 1995) was employed for calculating the secondary structural features of the antifreeze proteins and the result was presented in Table 5. Computational methods were also applied for determining disulphide bonds. Disulphide bonds are very essential in determining the functional linkage and the stability of a particular protein. The presence of SS bond and their bonding patterns were predicted by CYS_REC and What If server. CYS_REC (http://linux1.softberry.com/berry.phtml?topic) identified the position of a cystiene, total number of cystiene presented along with the most probable SS bond pairs in the protein sequences (Table 6). The later one What If involves the identification of SS bonds using the 3D structure of a protein.

Homology modeling and validation

Homology models of proteins are of great interest for planning and analyzing biological experiments when no experimental three dimensional structures are available. Many proteins are simply too large for NMR analysis and cannot be crystallized for X-ray diffraction. Protein modeling is the only way to obtain structural information if experimental techniques fail. Therefore, it is an obvious demand to bridge this 'structure knowledge gap' and computational methods for protein structure prediction have gained much interest in recent years (Schwede *et al*., 2003). The modeling of 3D structure of 2 antifreeze proteins were performed by three homology modeling programs Geno3D (Combet *et al*., 2002), Swiss-model (Arnold *et al*., 2006), CPH-models (Nielsen *et al*., 2010). Homology modeling of these two proteins was done by using a template structure from PDB (http://www.pdb.org/pdb/home/home.do) through BLASTP search (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The modeled 3D structures were evaluated using the online server Rampage, ProQ (Protein quality server) and ProSA. The structure validation of antifreeze proteins was performed by online PROCHECK (Laskowski *et al*., 1996) and What IF (Vriend, 1990).

RESULTS AND DISCUSSION

A total of 15 Antifreeze protein sequences of fishes were retrieved from SWISS-PROT and analyzed. The primary structure analysis was done and different parameters computed using EsPasy ProtParam tool was tabulated in table 3. The results of primary structure analysis suggest that, proteins from fishes are mostly hydrophobic and their hydrophobic nature is due to the presence of high

Table 3: Physicochemical properties of AFPs from different Fish varieties are computed using Expasy's ProtParam tool.

Accession Number	Length	M. wt.	P^I	(-) R	(+) R	EC	II	AI	GRAVY
P11920	19	1655.8	9.79	0	1	NIL	29.75	57.89	0.453
Q01758	175	19053.9	5.16	16	9	42105	33.3	76.46	0.171
Q92006	87	9408.4	7.91	6	7	1490	16.37	120.92	0.594
Q1AMQ4	86	9320.2	8.85	7	9	490	29.57	104.19	0.312
Q1AMR1	147	16364.8	4.85	16	9	42105	40.55	77.07	0.063
Q1AMQ2	88	9334.2	7.96	4	5	1490	14.19	120.8	0.588
P84493	147	16225.4	4.55	16	7	43595	33.75	69.73	0.057
Q1AMQ8	88	9470.4	9.36	4	7	1490	28.18	108.41	0.494
B1P0S1	218	19303.7	5.16	10	9	6990	13.03	97.02	1.026
Q1AMQ1	128	14377.5	4.8	16	11	1490	41.21	97.66	-0.218
Q1AMQ6	87	9398.4	7.89	6	7	1490	16.26	116.44	0.563
Q1AMR0	175	18993.8	5.16	16	9	42105	33.30	76.46	0.151
Q1AMR3	82	7767.8	6.00	5	5	5500	24.69	86.71	0.599
Q1AMQ9	88	9514.4	5.50	5	4	2980	17.06	121.70	0.659
Q1AMQ3	88	9408.3	7.98	4	5	1490	13.34	125.23	0.672

Legends: M. wt., P^I, (-) R, (+) R, EC, II, AI and GRAVY denotes Molecular weight, Isoelectric point, Positive R group, Negative R group, Extinction coefficient, Instability index, Aliphatic index and The grand average hydropathy.

non-polar residues. The presence of 11 (6.3%) Cys in Q01758 (Rainbow smelt), 11 (7.5%) Cys in Q1AMR1 (Atlantic herring), 11 (7.5%) Cys in P84493 (Japanese smelt) and 11 (6.3%) Cys in Q1AMR0 indicate the presence of disulphide bonds in corresponding Antifreeze protein. Moreover, the primary structure also suggests that the AFP P11920 has no aromatic residues (Phe, Trp and Tyr). The computed isolelectric point (p^I) will be useful because solubility is least at that p^I mobility in an electrofocusing system is zero. The isolelectric point (p^I) is the value at which the molecule carries no charges or the negative and positive charges are equal. The computed p^I value of AFPs which have p^I <7 indicates that these AFPs are acidic and p^I >7 indicate the basic nature of corresponding AFPs. The highest (9.79) and the lowest p^I value (4.55) was obtained from P11920 (Saffron cod) and P84493 (Japanese smelt) respectively, where the former one is basic and the later one is acidic in character. Most fish antifreeze proteins have basic character (according to retrieved protein sequences) with p^I value in average 6.73 respectively. For the purification of a particular protein by isoelectric focusing methods, the p^I value of this protein will be useful for developing buffer system. Extinction coefficient (EC) of AFPs were calculated by EsPasy protparam at 280nm wavelength is ranging from 1490 to 43595 M^{-1} cm^{-1} with respect to the concentration of Cys, Trp and Tyr. The high EC value of P84493, Q1AMR1, Q1AMR0 and Q01758 indicates presence of high concentration of Cys, Trp

Table 4: Transmembrane region identified by SOSUI server.

Accession no.	Transmembrane region (N terminal –C terminal)	Type	Length
B1P0S1	MALSLFTVGQFIFLFWTISITEA	PRIMARY	23
	ASKAAVTAADAAAAAATIAASAA	SECONDARY	23
	DTAAAAASAAAAAVASAAKALE	SECONDARY	22
	TAAAAAAATATTAAAAAAAKAT	SECONDARY	22
	AAVATAVSDAAATAATAAAVAAA	SECONDARY	23
	AAATAVSAAAAAAAAAIAFAAA	PRIMARY	22
Q1AMQ9	MKSVILTGLLFVLLCVDHMSSAN	PRIMARY	23
	ATQLIPINTALTLVMMTTRVIYP	SECONDARY	23
Q1AMR3	MALSLFTVGQLIFLFWTMRITEA	PRIMARY	23

Table 5: Calculated secondary structure features by SOPMA.

Accession Number	Secondary structure features			
	Alpha helix	Extended	Beta turn	Random coil
P11920	31.58%	0.00%	0.00%	68.42%
Q01758	24.57%	19.43%	5.14%	50.86%
Q92006	51.72%	14.94%	1.15%	32.18%
Q1AMQ4	69.77%	6.98%	2.33%	20.93%
Q1AMR1	28.57%	21.77%	3.40%	46.26%
Q1AMQ2	51.14%	12.50%	2.27%	34.09%
P84493	33.33%	18.37%	4.76%	43.54%
Q1AMQ8	45.45%	10.23%	5.68%	38.64%
B1P0S1	89.91%	4.59%	0.92%	4.59%
Q1AMQ1	97.66%	0.00%	1.56%	0.78%
Q1AMQ6	56.32%	10.34%	1.15%	32.18%
Q1AMR0	29.70%	17.14%	4.57%	48.57%
Q1AMR3	71.95%	10.98%	1.22%	15.85%
Q1AMQ9	45.45%	12.50%	5.68%	36.36%
Q1AMQ3	45.45%	9.09%	3.41%	42.05%

and Tyr. EsPasy protparam computes no EC value for P11920 due to the absence of Cys, Trp and Tyr. This indicates that this AFP cannot be analyzed using UV spectral methods. The computed EC values help in the quantitative study of protein-protein and protein-ligand interactions in solution. The instability index value of AFPs was calculated by EsPasy protparam which provides an estimation 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 (Guruprasud *et al*., 1990). The instability indexes of AFPs are ranging from 13.03 to 41.21. The highest instability index value was obtained from Q1AMQ1 (41.21) which is followed by Q1AMR1 (40.55), Q01758 (33.3), Q1AMR0 (33.3), and so on. Contrarily, the lowest instability index value was obtained from AFP B1P0S1 (13.03) of Winter flounder. The aliphatic index (AI) which is defined as 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. Aliphatic index of AFPs ranged from 57.89 (P11920) to 125.23 (Q1AMQ3) among sequences of different fish varieties. The lower thermal stability of P11920 and P84493 is indicative of a more flexible structure when compared to other AFPs. The very high aliphatic index of Q1AMQ3, Q1AMQ9, Q92006,

Table 6: Disulphide (SS) bond pattern of pairs predicted, by CYS_REC (using primary structure) and identified by what if (using 3D structure modeled).

Accession number	CYS_REC	What If
Q1AMR1	Cys 32- Cys 106	Cys 21- Cys 32
	Cys 49- Cys 142	Cys 49- Cys 142
	Cys 86- Cys 128	Cys 86- Cys 117
	Cys118- Cys 134	Cys 106- Cys 128
		Cys 118- Cys 134
P84493	Cys 21- Cys 117	Cys 21- Cys 32
	Cys 32- Cys 49	Cys 49- Cys 142
	Cys 86- Cys 128	Cys 86- Cys 117
	Cys 106- Cys 142	Cys 106- Cys 128
	Cys 118- Cys 134	Cys 118- Cys 134

Q1AMQ2, Q1AMQ6 and Q1AMQ8 infers that these AFPs may be stable for a wide range of temperature where all of them are type III Antifreeze protein. The Grand Average Hydropathy (GRAVY) value for a protein is calculated as a sum of hydropathy value of all amino acids, divided by the number of residues in the sequences. GRAVY index of all AFPs are ranging from -0.218 (Q1AMQ1) to 1.026 (B1P0S1) and infers that almost all fish antifreeze proteins analyzed in this study are hydrophobic. The very low GRAVY index of AFP Q1AMQ1 indicates the possibility of better interaction with water. Besides all other physicochemical characterization, functional characterization of antifreeze protein was also performed including transmembrane (TM) region identification, predict-tion of disulphide bonding pairs etc. The SOSUI server performed the identification transmembrane helices with their corresponding length and differentiates membrane proteins from stable proteins. The server SOSUI classifies B1P0S1,

Table 7: Ramachandran plot calculation and comparative analysis of models from Swiss-model, Geno3D and CPHmodels computed with PROCHECK program.

Server	Accession number	Rampage analysis		
		RFR	RAR	ROR
Swiss-model	Q1AMR1	93.6%	5.6%	0.8%
	P84493	92.6%	6.6%	0.8%
Geno3D	Q1AMR1	87.2%	12.0%	0.8%
	P84493	84.6%	12.2%	3.3%
CPHmodels	Q1AMR1	93.6%	6.4%	0.0%
	P84493	94.3%	4.9%	0.8%

Table 8: Protein 3D model of targets Q1AMR1 and P84493 from three different homology modeling server and validation parameter computed by ProQ and What If server.

Server	Accession number	Template (PDB) code	RMS Z score	ProQ	
				LG score	Maxsub
Swiss-model	Q1AMR1	2PY2_A	0.909	2.491	0.288
	P84493	2PY2_A	0.924	2.283	0.304
Geno3D	Q1AMR1	2PY2_A	0.467	2.232	0.253
	P84493	2PY2_A	0.471	2.246	0.303
CPHmodels	Q1AMR1	2PY2_A	0.927	1.828	0.214
	P84493	2PY2_A	0.927	2.377	0.306

Q1AMQ9 and Q1AMR3 as membrane protein and others as soluble proteins. This antifreeze membrane proteins B1P0S1 (Winter flounder) contains 6 TM helices, Q1AMQ9 and Q1AMR3 has 2 and 1 TM helix respectively. The TM helices and their length were tabulated in table 4. Hydrophobicity of these AFPs was also computed based on Kyte Dolittle hydrophobicity index by ProtScale (http://expasy.org/tools/protscale.html) and TMpred. The secondary structures of AFPs were predicted by SOPMA (self optimized prediction method with alignment) which correctly predicts 69.5% of amino acids for a state description of the secondary structure prediction (Geourgon and Deleage 1995). This secondary structure indicates whether a given amino acid lies in a helix, strand or coil. Calculated secondary structure features were tabulated in table 5. This result revealed that random coils dominated among secondary structure features followed by alpha helix, extended strands and beta turns for all sequences while all other secondary structure features such as 3_{10} helix, Pi helix, Ambiguous states, Bend region and Beta bridge were not found. Alpha helix is the dominating secondary structure feature in Fish AFPs. The secondary structure were predicted by using default parameters (Windows width: 17, similarity threshold: 8, and number of states: 4). The tool CYS_REC identifies the presence of S-S bonds and 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

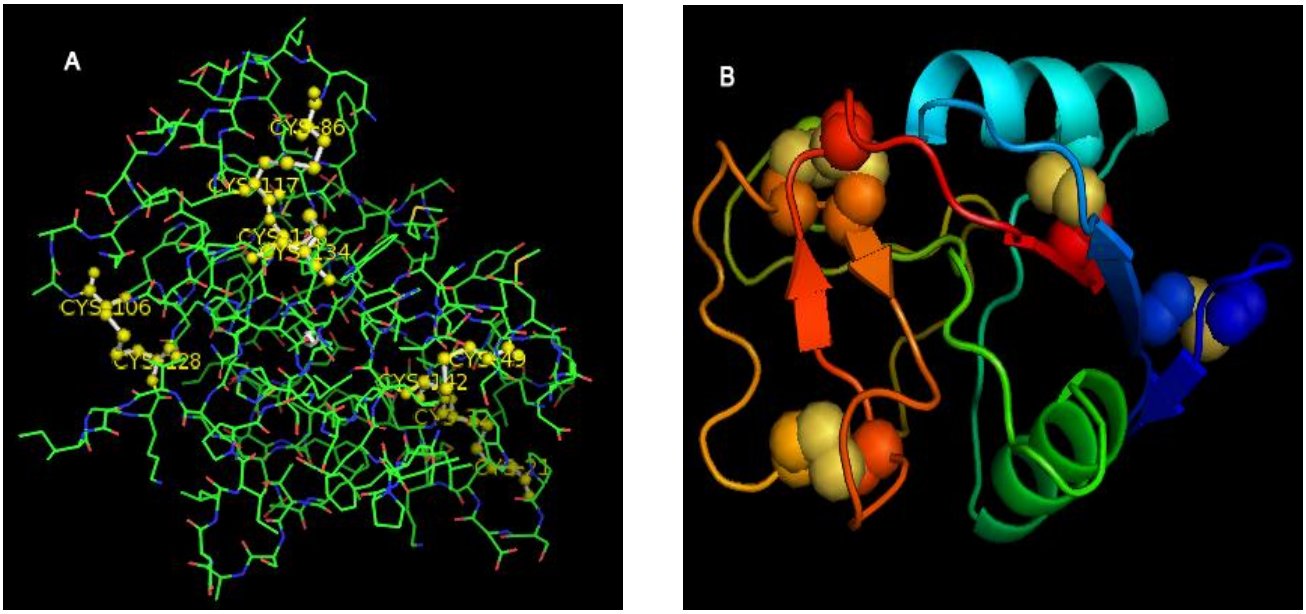

Figure 1: PyMol representation (wireframe diagram and strands) of the homology modeled 3D structure of fish antifreeze protein (A) Q1AMR1 (Atlantic herrings) Cystiene residues are shown as ball and stick models (Yellow). (B) P84493 (Japanese smelt) disulphide bonds are shown as spheres models (Yellow).

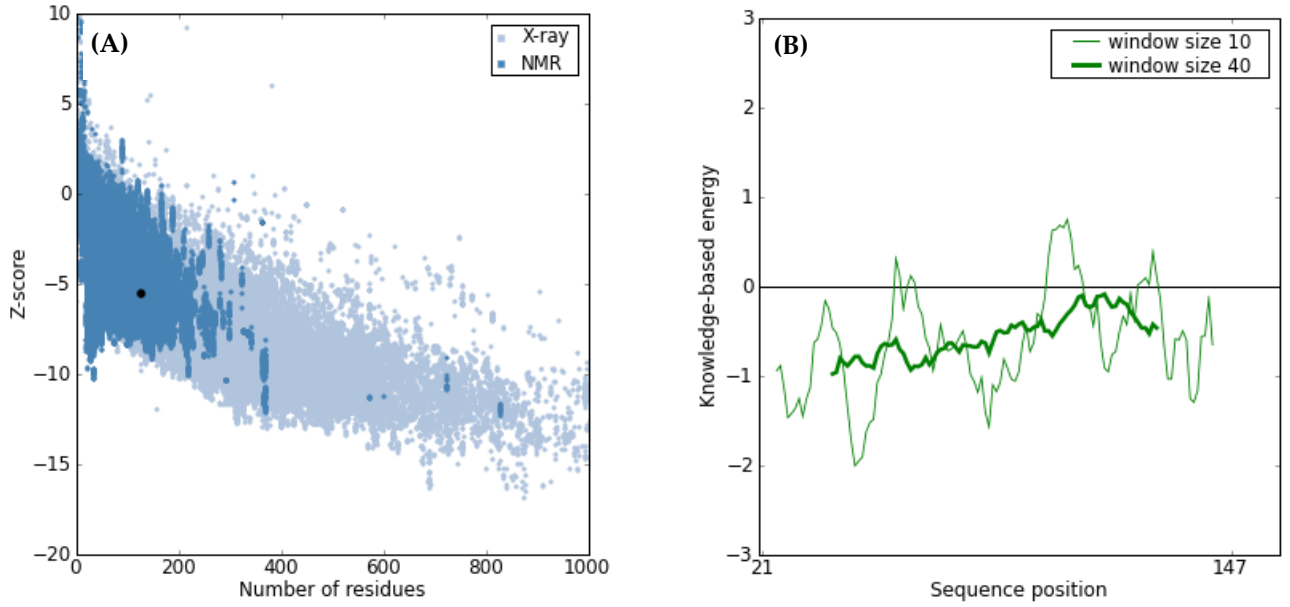

Figure 2: ProSA-web service analysis of AFP Q1AMR1. (A) ProSA-web z-scores of all protein chains in PDB determined by X-ray crystallography (light blue) or NMR spectroscopy (dark blue) with respect to their length. The z-scores of Q1AMR1 highlighted as large dot. (B) Energy plot of Q1AMR1.

were identified from 3D structure by "What If" in the AFPs Q1AMR1 and P84493 are shown in table 6.

Homology modeling and model validation

Three-dimensional (3D) protein structures provide valuable insights into the molecular basis of protein function, allowing an effective design of experiments. Homology models of proteins are of great interest for planning and analyzing biological experiments when no experimental three dimensional structures are available. Now a day, 3D structure of protein can be predicted from amino acid sequences by different web based homology modeling servers at different level of complexity. During evolution, the structure is more stable and changes much slower than the associated sequence, so that similar sequences adopt practically identical structures and distantly related sequences still fold into similar structures (Chothia and Lesk 1986). The modeling of 3D structure of protein was performed by three homology modeling program Geno3D, Swiss model and CPHmodels. Two antifreeze proteins Q1AMR1 (Atlantic herring), P84493 (Japanese smelt) are considered for homology modeling based on PDB template selected from the hits obtained through the BLASTP analysis. The stereo chemical quality of the predicted models and accuracy of the protein model was verified after the refinement process using Ramchandran Map calculation computed with PROCHECK program (Laskowski *et al.*, 1993). PROCHECK suite of a program for assessing the stereo chemical quality of a given protein structure and to measure how normal or conversely how unusual, the geometry of the residues in a given protein model is as compared with stereo chemical parameters derived from well refined high resolution structure. The result revealed that, the proteins Q1AMR1 and P84493 modeled by Swiss model homology modeling server has average maximum residues in favored region (RFR) which are about 93.6% and 92.6% respectively. A comparison of the results obtained from three different modeling server in table 7 shows that the models generated by Swiss

Table 9: Criteria for a good (model) 3D structure.

Rampage percentage of residues in favored region	RMS Z score	ProQ		Quality of the model
		LG score	Maxsub	
98	1	>1·5	>0·1	Fairly good model
		>2·5	>0·5	Very good model
		>4	>0·8	Extremely good model

model was more acceptable in comparison with others. The modeled structure of antifreeze proteins were also validated by other model verification servers What If and Protein Quality Server (ProQ), each of which validates protein models based on different validation parameters. Two quality measures, LG score and MaxSub of three models from each modeling server are predicted by ProQ and listed with RMS Z score in table 8. Criteria for a good 3D model are given in table 9. The result revealed RMS Z score, LG score, MaxSub and other criterions suggesting good model quality except the models generated by Geno3D. The cystienes and disulphide bonds identified using 3D structure of AFPs Q1AMR1 and P84493 are shown in Figure 1. Some S-S bonding pairs predicted by CYS_REC are not correlating with the S-S bond positions identified using 'What If'. We speculate that, S-S bonds predicted from 3D structure might be correct and more reliable than the S-S bonds identified from the primary structure. ProSA was used to check three dimensional models of AFPs for potential errors. The program displays two quality measures of the input structure; z-score and a plot of its residue energies. The z-score indicates overall model quality and measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. As shown in Figure 2(A) the Z-score for AFPs are also well within the range of scores typically found for proteins of similar size indicating a highly reliable structure. The energy plot shows the local model quality by plotting energies as a function of amino acid sequence position. In general, positive values correspond to problematic or erroneous parts of a model. Figure 2(B) displays a comparable energy plot for both the target and template structures.

REFERENCES

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling, Bioinformatics 22: 195-201. http://dx.doi.org/10.1093/bioinformatics/bti770 PMid:16301204

Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003, Nucleic Acids Res, 31:365-370. http://dx.doi.org/10.1093/nar/gkg095 PMid:12520024 PMCid:165542

Combet C, Jambon M, Deleage G, Geourjon C (2002) Geno3D: Automatic comparative molecular modelling of protein, Bioinformatics 18: 213-214. http://dx.doi.org/10.1093/bioinformatics/18.1.213 PMid:11836238

Davies PL and Hew CL (1990) Biochemistry of fish antifreeze proteins, FASEB J. 4: 2460-2468. PMid:2185972

Davies PL and Sykes BD (1997) Antifreeze proteins, Curr. Opin. Struct. Biol. 7: 828–834. http://dx.doi.org/10.1016/S0959-440X(97)80154-6

Deng C, Cheng CHC, Ye H, He X, and Chen L (2010) Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict, PNAS 2010 107: 21593-21598. http://dx.doi.org/10.1073/pnas.1007883107 PMid:21115821 PMCid:3003108

Fletcher GL, Goddard SV, Wu YL (1999) Antifreeze proteins and their genes: From basic research to business opportunity, ChemTech 29: 17–28.

Fletcher GL, Hew CL and Davies PL (2001). Antifreeze proteins of teleost fishes, Annu Rev Physiol 63: 359–390. http://dx.doi.org/10.1146/annurev.physiol.63.1.359 PMid:11181960

Food Standards Australia New Zealand (FSANZ) (2006) Ice structuring protein as a processing aid in ice cream and edible ices; a safety assessment. Technical report series, 42.

Gasteiger E (2005) Protein Identifi cation and Analysis Tools on the ExPASy Server. In: John M. Walker ed, The Proteomics Protocols Handbook, Humana Press, 571-607.

Geourjon C, Deléage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, Comput Appl Biosci 11: 681-684. PMid:8808585

Gilbert JA, Davies PL, Laybourn-Parry J (2005) A hyperactive, Ca2+ -dependent antifreeze protein in an Antarctic bacterium, FEMS Microbiol Lett 245: 67–72. http://dx.doi.org/10.1016/j.femsle.2005.02.022 PMid:15796981

Gilbert JA, Hill PJ, Dodd CER and Laybourn-Parry J (2004) Demonstration of antifreeze protein activity in Antarctic lake bacteria, Microbiology, 150: 171–180. http://dx.doi.org/10.1099/mic.0.26610-0 PMid:14702410

Gill SC, Von Hippel PH (1989) Extinction coefficient, Anal Biochem 182: 319- 328. http://dx.doi.org/10.1016/0003-2697(89)90602-7

Graether SP and Sykes BD (2004) Cold survival in freeze-intolerant insects: the structure and function of beta-helical antifreeze proteins, Eur. J. Biochem. 271: 3285–3296 http://dx.doi.org/10.1111/j.1432-1033.2004.04256.x PMid:15291806

Griffith M and Yaish MW (2004) Antifreeze proteins in overwintering plants: a tale of two activities, Trends Plant Sci. 9: 399–405. http://dx.doi.org/10.1016/j.tplants.2004.06.007 PMid:15358271

Guruprasad K, Reddy BVP, Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence, Prot Eng 4: 155-164. http://dx.doi.org/10.1093/protein/4.2.155

Hoshino T, Kiriaki M and Nakajima T (2003) Novel thermal hysteresis proteins from low temperature basidiomycete, Coprinus psychromorbidus, Cryo Letters. 24: 135–142. PMid:12908023

Ikai AJ (1980) Thermo stability and aliphatic index of globular proteins. J Biochem 88: 1895-1898. PMid:7462208

Jorov A., Zhorov BS and Yang DS (2004) Theoretical study of interaction of winter flounder antifreeze protein with ice, Protein Sci. 13: 1524-1537. http://dx.doi.org/10.1110/ps.04641104 PMid:15152087 PMCid:2279984

Kawahara H., Nakano Y, Omiya K, Muryoi N, Nishikawa J and Obata H (2004) Production of two types of ice-crystal controlling proteins in Antarctic bacterium, J. Biosci. Bioeng. 98: 220–223. PMid:16233695

Kyte J, Doolottle RF (1982) A simple method for displaying the hydropathic character of a protein, J Mol Biol 157: 105-132. http://dx.doi.org/10.1016/0022-2836(82)90515-0

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, J Biomol NMR 8: 477-486. http://dx.doi.org/10.1007/BF00228148 PMid:9008363

Lin FH, Graham LA, Campbell RL, and Davies PL (2007) Structural Modeling of Snow Flea Antifreeze Protein, Biophysical Journal, 92: 1717–1723 http://dx.doi.org/10.1529/biophysj.106.093435 PMid:17158562 PMCid:1796811

Liu Y, Li Z, Lin Q, Kosinski J, Seetharaman J, (2007) Structure and Evolutionary Origin of Ca2+-Dependent Herring Type II Antifreeze Protein, PLoS ONE 2(6): e548. http://dx.doi.org/10.1371/journal.pone.0000548 PMid:17579720 PMCid:1891086

Madura JD, Baran K, and Wierzbicki A (2000) Molecular recognition and binding of thermal hysteresis proteins to ice, J. Mol. Recognit. 13:101–113. [DOI]

Nielsen M, Lundegaard C, Lund O and Petersen TN (2010) CPHmodels-3.0—remote homology modeling using structure-guided sequence profiles, Nucleic Acids Research, 38: W576-W581. http://dx.doi.org/10.1093/nar/gkq535 PMid:20542909 PMCid:2896139

Raymond JA and DeVries AL (1977) Adsorption inhibition as a mechanism of freezing resistance in polar fishes, Proc Natl Acad Sci; 74(6): 2589–2593. http://dx.doi.org/10.1073/pnas.74.6.2589

Sandve SR, Rudi H, Asp T and Rognli OA (2008) Tracking the evolution of a cold stress associated gene family in cold tolerant grasses, BMC Evolutionary Biology, 8: 245. http://dx.doi.org/10.1186/1471-2148-8-245 PMid:18775065 PMCid:2542378

Schwede T, Kopp J, Guex N and Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server, Nucleic Acids Research, 13: 3381-3385. http://dx.doi.org/10.1093/nar/gkg520 PMid:12824332 PMCid:168927

Tobias JW, Shrader TE, Rocap G and Varshavsky A (1991) The N-end rule in bacteria, Science 254 (5036): 1374-1377. http://dx.doi.org/10.1126/science.1962196 PMid:1962196

Vriend G (1990) WHAT IF: A molecular modeling and drug design program, J Mol Graph 8: 52-56. http://dx.doi.org/10.1016/0263-7855(90)80070-V

Wang R, Zhang P, Gong Z, Hew CL (1995) Expression of the antifreeze protein gene in transgenic goldfish (Carassius auratus) and its implication in cold adaptation, Mol Mar Biol Biotechnol; 4: 20–26. PMid:7749462