J. Chem. Sci., Vol. 119, No. 5, September 2007, pp. 571–579.

In silico characterization of antifreeze proteins using computational tools and servers

K SIVAKUMAR,[1,*] S BALAJI[1] and GANGARADHAKRISHNAN[2]

[1]Department of Chemistry, Sri Chandrasekharendra Saraswathi Viswa Maha Vidyalaya (Deemed University), Enathur, Kanchipuram 631 561

[2]EXCEL and Polymer Science Labs, Central Leather Research Institute, Adyar, Chennai 600 020

e-mail: shivamk25@yahoo.co.in

MS received 13 June 2007; accepted 9 August 2007

Abstract. In this paper, seventeen different fish Antifreeze Proteins (AFPs) retrieved from Swiss-Prot database are analysed and characterized using *In silico* tools. Primary structure analysis shows that most of the AFPs are hydrophobic in nature due to the high content of non-polar residues. The presence of 11 cysteines in the rainbow smelt fish and sea raven fish AFPs infer that these proteins may form disulphide (SS) bonds, which are regarded as a positive factor for stability. The aliphatic index computed by ExPasy's ProtParam infers that AFPs may be stable for a wide range of temperature. Secondary structure analysis shows that most of the fish AFPs have predominant α-helical structures and rest of the AFPs have mixed secondary structure. The very high coil structural content of rainbow smelt fish and sea raven fish AFPs are due to the rich content of more flexible glycine and hydrophobic proline amino acids. Proline has a special property of creating kinks in polypetide chains and disrupting ordered secondary structure. SOSUI server predicts one transmembrane region in winter flounder fish and atlantic cod and two transmembrane regions in yellowtail flounder fish AFP. The predicted transmembrane regions were visualized and analysed using helical wheel plots generated by EMBOSS pepwheel tool. The presence of disulphide (SS) bonds in the AFPs Q01758 and P05140 are predicted by CYS_REC tool and also identified from the three-dimensional structure using Rasmol tool. The disulphide bonds identified from the three-dimensional structure using the Rasmol tool might be correct as the evaluation parameters are within the acceptable limits for the modelled 3D structures.

Keywords. Antifreeze proteins; computational analysis; disulphide bridges; homology modelling; proteomics tools.

1. Introduction

Computational packages and online servers are the current tools used in the protein sequence analysis and characterization.[1] The physicochemical and the structural properties of the proteins are well understood with the use of computational tools. Today, number of computational tools has been developed for making predictions regarding the identification and structure prediction of proteins. The statistics about a protein sequence such as number of amino acid, sequence length, and the physico-chemical properties of a proteins such as molecular weight, atomic composition, extinction coefficient, GRAVY, aliphatic index, instability index, etc. can be computed by computational tools for the prediction and characterization of protein structure. The amino acid sequence provides most of the information required for determining and characterizing the molecule's function, physical and chemical properties. Sequence analysis and physicochemical characterization of proteins using biocomputation tools have been done by many researches and reported.[2–8] Antifreeze Proteins (AFPs) were first identified in fishes. Antifreeze proteins resist ice crystal growth and prevent cellular damage in the organisms due to freezing. AFP molecules have a strong affinity for ice due to their structure. AFPs protect the organism from freezing at temperature below 1°C by interacting with small ice crystals and inhibit their growth through an adsorption–inhibition mechanism.[9] Many researchers have purified and analysed macromolecular antifreeze proteins from a number of plants, fishes and insects. To date researches have identified five different AFPs from fish and they are classified as Antifreeze Glycoproteins and Antifreeze proteins Type I, Type II,

*For correspondence

Type III and Type IV based on their properties, molecular weight and structure.[10] AFPs are highly useful in the preservation techniques because of their recrystallization inhibition property.[11] AFPs have potential applications in agriculture for protecting crops from freezing, in maintaining the texture in frozen foods and for producing cold-hardy plants using transgenic technology. AFPs are used in the cryosurgery for the low temperature preservation of cells, tissues and organs.[12] Chao *et al*[13] have reported the relative efficacy of AFP types I, II and III in protecting the red blood cells. Numerous structure and function studies have been reported from time to time from all over the world.[14–18] However, physico-chemical characterization of antifreeze protein has not been done so far. In this paper, we report the *In silico* analysis and characterization studies on 17 AFPs of various fishes.

2. Materials and methods

2.1 *Antifreeze protein sequences*

Antifreeze protein sequences were retrieved from the manually curated public protein database Swiss-Prot.[19] Swiss-Prot is scanned for the key word antifreeze. The search result yielded 39 antifreeze protein sequences of 17 fishes. From this, we have retrieved 17 different fish AFPs (i.e. one antifreeze protein is chosen from each type of fish) by random selection and have organized a non-redundant data set (table 1). The AFPs were retrieved in FASTA format and used for analysis.

2.2 *Computational tools and servers*

The amino acid composition (table 2) of AFP sequences were computed using the tool CLC *free Workbench.*[20] Percentages of hydrophobic and hydrophilic residues were calculated from the primary structure analysis results and tabulated in table 3. The physico-chemical parameters, theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient,[21] half-life,[22–25] instability index,[26] aliphatic index[27] and grand average hydrophathy[28] (GRAVY) were computed using the Expasy's ProtParam (http://us.expasy.org/tools/protparam.html) prediction server and tabulated in table 4. The tools SOPM, SOPMA[29] and Secondary Structural Content Prediction (SSCP method-I) server[30] were used for the secondary structure prediction. The SOSUI[31] server performed the identification of transmembrane regions (table 5). The predicted transmembrane helices were visualized and analysed using helical wheel plots (figure 2) generated by the program Pepwheel[32] included in the EMBOSS 2.7 suite. The presence of disulphide bridges (SS bonds) in AFPs Q01758 and P05140 is predicted by two methods. The first method involves the prediction of SS bonds using the primary structure (protein sequence data) by the tool CYS_REC.[33] CYS_REC identifies the positions of cysteines, total number of cysteines present and predicts the most probable SS bond pattern of pairs in the protein sequence. The second method involves the visualization and identification of SS bonds using the three-dimensional

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

Accession number	Sequence description	Organism
P20617	Antifreeze peptide GS-8	Grubby sculpin (*Myoxocephalus aenaeus*)
P04368	Antifreeze peptide SS-8	Shorthorn sculpin (*Myoxocephalus scorpius*)
P80961	Antifreeze protein LS-12	Longhorn sculpin (*Myoxocephalus octodecimspinosis*)
P24028	Antifreeze protein LP	Canadian eelpout (*Lycodes polaris*)
P12101	Antifreeze peptide AB2	Antarctic eelpout (*Pachycara brachycephalum*)
P35751	Antifreeze peptide RD1	Antarctic eelpout (*Rhigophila dearborni*)
P07457	Antifreeze protein SP1-C	Ocean pout (*Macrozoarces americanus*)
P12417	Antifreeze protein type III	Atlantic wolfish (*Anarhichas lupus*)
Q01758	Type II antifreeze protein	Rainbow smelt (*Osmerus mordax*)
P05140	Type II antifreeze protein	Sea raven (*Hemitripterus americanus*)
P04002	Antifreeze protein A	Winter flouonder (*Pseudopleuronectes americanus*)
P09031	Antifreeze protein	Yellowtail flounder (*Limanda ferruginea*)
Q8JI37	Type-4 ice-structuring protein	Japanese flounder (*Paralichthys olivaceus*)
Q56TU0	Type-4 ice-structuring protein precursor	Atlantic cod (*Gadus morhua*)
P11920	Antifreeze glycoprotein 7R	Saffron cod (*Eleginus gracilis*)
P02732	Antifreeze glycoprotein 3	Bald rockcod (*Pagothenia borchgrevinki*)
P24856	Antifreeze glycopeptide polyprotein	Black rockcod (*Notothenia coriiceps neglecta*)

Table 2. Amino acid composition (in %) of antifreeze proteins computed using CLC free WorkBench tool.

Amino acids	Accession number																
	P20617	P04368	P80961	P24028	P12101	P35751	P07457	P12417	Q01758	P05140	P04002	P09031	Q8JI37	Q56TU0	P11920	P02732	P24856
Ala	60	60	11·72	12·12	7·94	9·38	8·05	11·36	10·86	13·5	46·34	52·58	14·52	9·6	57·89	67·74	49·75
Cys	0	0	0	0	0	0	1·15	1·14	6·29	6·75	0	0	0	0	0	0	0·25
Asp	5	2·22	4·69	3·03	3·17	1·56	1·15	2·27	5·14	5·52	3·66	6·19	3·23	3·2	0	0	0
Glu	2·5	2·22	7·81	3·03	9·52	6·25	2·3	2·27	4	3·07	2·44	2·06	8·87	11·2	0	0	0
Phe	0	0	3·91	0	0	0	2·3	2·27	4	1·84	3·66	3·09	5·65	2·4	0	0	6·46
Gly	2·5	2·22	2·34	6·06	3·17	4·69	6·9	4·55	6·29	6·75	2·44	2·06	0·81	2·4	0	0	0·13
His	0	0	0·78	0	0	0	1·15	1·14	3·43	2·45	0	0	0·81	0·8	0	0	0·51
Ile	2·5	2·22	7·81	7·58	6·35	7·81	6·9	6·82	4·57	3·07	2·44	2·06	4·03	5·6	0	0	2·53
Lys	7·5	8·89	7·03	3·03	6·35	6·25	6·9	6·82	4	3·07	2·44	3·09	6·45	7·2	0	0	0
Leu	5	4·44	10·16	10·61	9·52	9·38	9·2	9·09	8	7·98	7·32	5·15	10·48	12·8	0	0	4·68
Met	2·5	2·22	4·69	6·06	7·94	9·38	8·05	7·95	4·57	5·52	2·44	1·03	4·03	4·8	0	0	0
Asn	0	2·22	3·91	6·06	6·35	7·81	4·6	4·55	4	3·07	2·44	1·03	5·65	0	0	0	2·66
Pro	2·5	2·22	2·34	9·09	7·94	9·38	6·9	5·68	5·14	5·52	6·1	4·12	2·42	4	10·53	0	9·24
Gln	2·5	2·22	14·84	4·55	3·17	3·13	5·75	5·68	2·29	4·91	1·22	1·03	12·9	12	0	0	0
Arg	2·5	2·22	1·56	4·55	3·17	1·56	0	2·27	1·14	1·84	2·44	2·06	2·42	1·6	5·26	0	0·13
Ser	0	2·22	3·13	1·52	3·17	1·56	4·6	6·82	6·29	6·75	3·66	2·06	3·23	5·6	0	0	0
Thr	5	4·44	7·03	7·58	6·35	7·81	10·34	6·82	9·14	8·59	8·54	8·25	8·06	8	26·32	32·26	23·54
Val	0	0	5·47	13·64	12·7	12·5	12·64	11·36	5·71	4·29	1·22	3·09	5·65	7·2	0	0	0·13
Trp	0	0	0	0	0	0	0	0	4	4·29	1·22	1·03	0	0	0	0	0
Tyr	0	0	0·78	1·52	3·17	1·56	1·15	1·14	1·14	1·23	0	0	0·81	1·6	0	0	0

Table 3. Hydrophilic and hydrophobic residues content.

Accession number	Percentage of hydrophobic residues	Percentage of hydrophilic residues	Net hydrophobic residues content
P20617	70	25	Very high
P04368	68·88	26·65	Very high
P80961	43·76	51·56	Low
P24028	50·01	34·87	High
P12101	44·45	44·42	–
P35751	48·45	37·49	High
P07457	56·32	43·67	High
P12417	48·85	39·78	High
Q01758	41·71	40·57	–
P05140	40·49	40·5	–
P04002	64·64	26·84	Very high
P09031	68·03	25·77	Very high
Q8JI37	45·96	54·03	Low
Q56TU0	46·4	53·6	Low
P11920	57·89	42·10	High
P02732	67·74	32·25	High
P24856	63·92	36·07	High

Table 4. Parameters computed using Expasy's ProtParam tool.

Accession number	Sequence length	M. wt	pI	– R	+ R	EC	II	AI	GRAVY
P20617	40	3579	8·25	3	4	Nil	10·56	89·25	0·59
P04368	45	4006·5	10	2	5	Nil	7·16	86	0·54
P80961	128	14377·5	4·8	16	11	1280	41·21	97·66	-0·218
P24028	66	6982·3	8·5	4	5	1280	26·31	122·58	0·489
P12101	63	7001·2	4·96	8	6	2560	24·2	106·67	0·056
P35751	64	6906·3	6·39	5	5	1280	17·25	112·66	0·353
P07457	87	9229	9·36	3	6	1490	21·29	107·47	0·531
P12417	88	9430·3	9·65	4	8	1280	30·24	106·36	0·43
Q01758	175	19053·9	5·16	16	9	42990	33·3	76·46	0·171
P05140	163	17509	4·93	14	8	42990	36·05	69·02	0·045
P04002	82	7710·7	4·86	5	4	5690	24·69	87·93	0·668
P09031	97	8864·9	4·41	8	5	5690	10·56	89·69	0·739
Q8JI37	124	12946	4·96	15	11	1490	34·42	87·50	–0·281
Q56TU0	125	13991·2	4·68	18	11	2980	47·4	102·24	–0·126
P11920	19	1655·8	9·79	0	1	Nil	29·75	57·89	0·453
P02732	31	2521·7	5·57	0	0	Nil	9·68	67·74	0·994
P24856	790	71266·5	8·08	0	1	125	24·41	78·25	0·952

M. wt., Molecular weight; pI, Isoelectric point; –R, Number of negative residues; +R, Number of positive residues; EC, Extinction coefficient at 280 nm; II, Instability index; AI, Aliphatic index; GRAVY, Grand Average Hydropathy.

structure of protein (3D coordinates data). The 3D structure of AFPs Q01758 and P05140 were generated by homology modelling using Esypred[34] server. The similar 3D structures (for the AFPs Q01758 and P05140 sequences) in the Protein Data bank (www.rscb.org) were identified by the BLASTP analysis (http://www.ncbi.nlm.nih.gov:80/BLAST/). The modelled 3D structures were evaluated using the online servers Rampage,[35] ProQ[36] (Protein Quality server) and CE[37] (Combinatorial Extension). The tool Rasmol (http://openrasmol.org/) is used to visualize the modelled 3D structures and to identify the SS bonds. The three-dimensional structures of AFPs Q01758 and P05140 modelled using the PDB template 2AFP_A are shown in figures 3 and 4 respectively. The five most probable SS bond pattern of

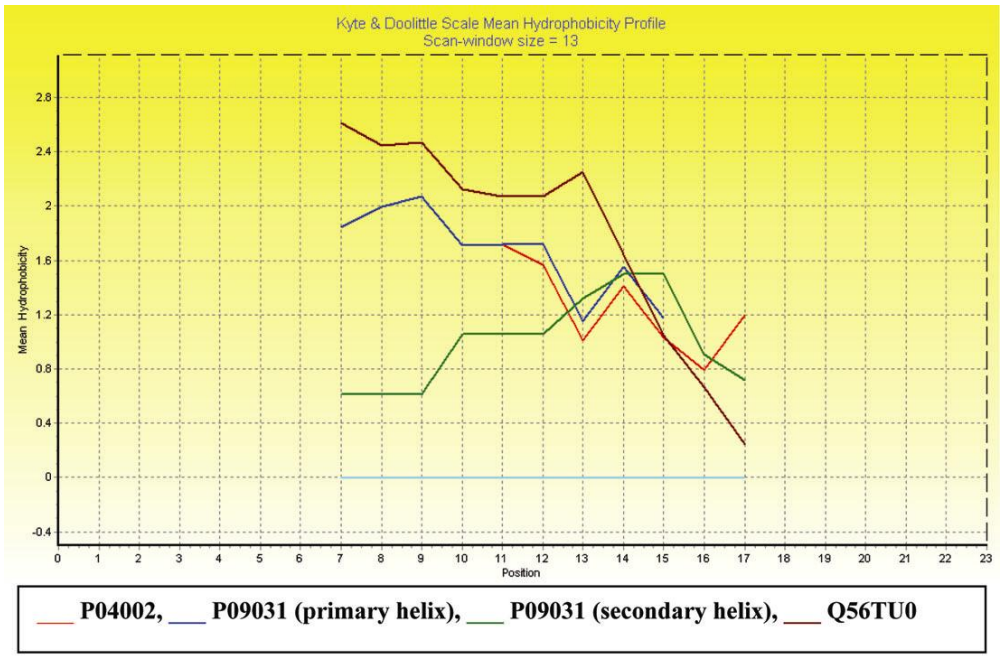

Figure 1. Kyte and Doolittle mean hydrophobicity profile computed for the transmembrane regions of AFPs P04002, P09031 (primary and secondary helices) and Q56TU0.

Table 5. Transmembrane regions identified by SOSUI server.

Accession number	Transmembrane region	Type	Length
P04002	MALSLFTVGQLIFLFWTMRITEA	Primary	23
P09031	MALSLFTVGQLIFLFWTLRIT	Primary	21
	AAKAAPAAVADPAAAAAAAVADT	Secondary	23
Q56TU0	YTLIAAIVVLALAQGTLAVEQSP	Primary	23

Table 6. Disulphide (SS) bond pattern of pairs predicted, by CYS_REC (using primary structure) and identified by Rasmol (using 3D structure modelled).

Accession number	CYS_REC	RasMol
Q01758	Cys38-Cys49	Cys38-Cys49
	Cys66-Cys135	Cys66-Cys159
	Cys103-Cys145	Cys135-Cys151
	Cys123-Cys134	Cys103-Cys134
	Cys151-Cys159	Cys123-Cys145
P05140	Cys41-Cys52	Cys41-Cys52
	Cys69-Cys159	Cys69-Cys159
	Cys103-Cys134	Cys103-Cys134
	Cys123-Cys145	Cys123-Cys145
	Cys135-Cys151	Cys135-Cys151

pairs predicted by CYS_REC tool and the positions of SS bonds identified using Rasmol tool in the AFPs Q01758 and P05140 are shown in table 6.

3. Results and discussion

The results of primary structure analysis suggest that most of the AFPs are hydrophobic in nature due to the presence of high non-polar residues content (tables 2 and 3). The presence of 11 Cys residues in AFPs Q01758 (6·29% of Cys) (rainbow smelt fish) and P05140 (6·75% of Cys) (sea raven fish) indicates the presence of disulphide bridges (SS bonds) in these AFPs. Moreover, the primary structure analysis suggests that the AFPs P20617, P04368, P11920 and P02732 have no aromatic residues (Tyr, Phe and Trp). The average molecular weight of AFPs calculated is 12766 Da. Isoelectric point (pI) is the pH at which the surface of protein is covered with charge but net charge of the protein is zero. At pI proteins are stable and compact. The computed pI value of P80961, P12101, P35751, Q01758, P05140, P04002, P09031,

Q8JI37 and Q56TU0 (pI < 7) indicates that these AFPs are acidic and the pI of P20617, P04368, P24028, P12417, P07457, P11920 and P24856 (pI > 7) reveals that these are basic in character. The computed isolelectric point (pI) will be useful for developing buffer systems for purification by isoelectric focusing method. Although Expasy's ProtParam computes the extinction coefficient for a range of (276, 278, 279, 280 and 282 nm) wavelength, 280 nm is favoured because proteins absorb strongly there while other substances commonly in protein solutions do not. Extinction coefficient of AFPs at 280 nm is ranging from 1280 to 42990 M^{-1} cm^{-1} with respect to the concentration of Cys, Trp and Tyr. The high extinction coefficient of Q01758 and P05140 indicates presence of high concentration of Cys, Trp and Tyr. Expasy's ProtParam computes no value for P20617, P04368, P11920 and P02732 because it has no Cys, Trp or Tyr. This indicates that these AFPs cannot be analysed using UV spectral methods. The computed protein concentration and extinction coefficients help in the quantitative study of protein–protein and protein–ligand interactions in solution. The biocomputed half-life of most of the AFPs is greater than 20 h. The half-life is only 4·4 h for P11920, P02732 and 3 min for AFPs P24028 and P35751. On the basis of instability index Expasy's ProtParam classifies the P80961 (Longhorn Sculphin fish) and Q56TU0 (Atlantic cod) AFPs as unstable (Instability index > 40) and other AFPs as stable (Instability index < 40). 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. The lower thermal stability of P05140, P11920 and P02732 is indicative of a more flexible structure when compared to other AFPs (table 4). The very high aliphatic index of all AFPs infers that AFPs may be stable for a wide range of temperature. Grand Average hydropathy (GRAVY) Index of AFPs are ranging from –0·1 to 0·9. The very low GRAVY index of AFPs P80961, Q8JI37 and Q56TU0 infers that these AFPs could result in a better interaction with water. The secondary structure predicted with the help of programs SOPM and SOPMA (data not shown) infers that the AFPs P20617, P04368, P80961 (Sculphin fishes),

Table 7. PDB templates (first 2 hits with maximum % identity) obtained using BLASTP search against the Protein Data Bank.

Accession number	PDB code
Q01758	2AFP_A 1XAR_B
P05140	2AFP_A 1QDD_A

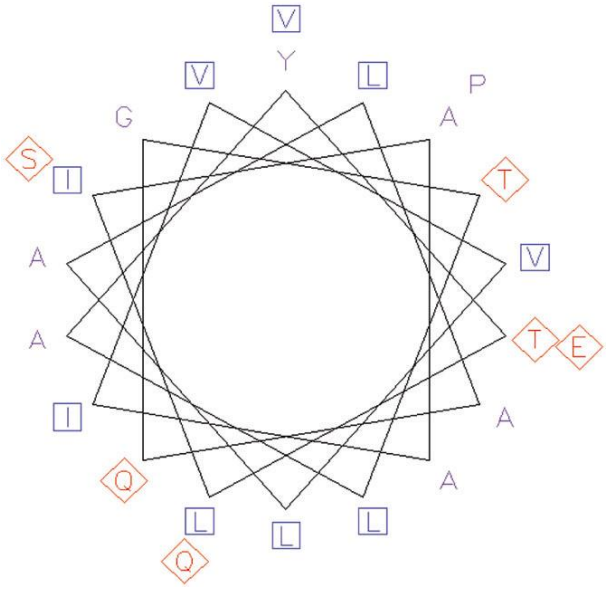

Figure 2. Helical wheel representation of predicted helix of Q56TU0 (Atlantic cod fish) AFP. Hydrophobic residues (V, L, I) are represented as blue squares and violet letters (A, G, P, Y), polar residues (E, Q, S, T) as red diamonds.

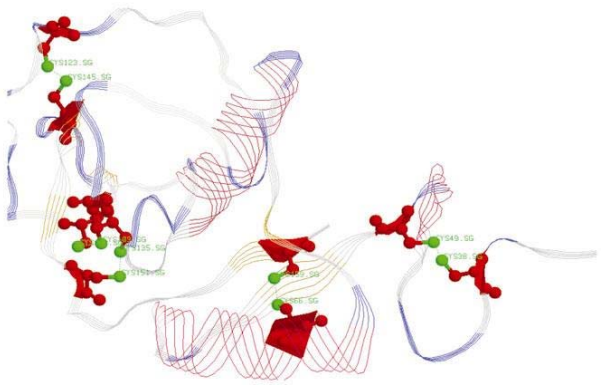

Figure 3. RasMol (strands) representation of the homology modelled 3D structure of antifreeze protein Q01758 (using PDB template 2AFP_A). The 10 cysteines are shown as ball and stick models (red). The sulphur atoms present in cysteines and the SS bonds (dotted lines) are shown in green colour. One unpaired cysteine is not shown.

Table 8. Validation parameters computed for the built 3D structures of targets Q01758 and P05140.

Target	Template (PDB) codes	RamPage Percentage of residues in favoured region	CE RMSD (Å)	ProQ	
				LG Score	Maxsub
Q01758	2AFP_A	82	0·5	1·464	0·245
	1XAR_B	65	2·0	0·9	0·063
P05140	2AFP_A	81·1	0·5	1·604	0·2
	1QDD_A	52	1·8	1·3	0·12

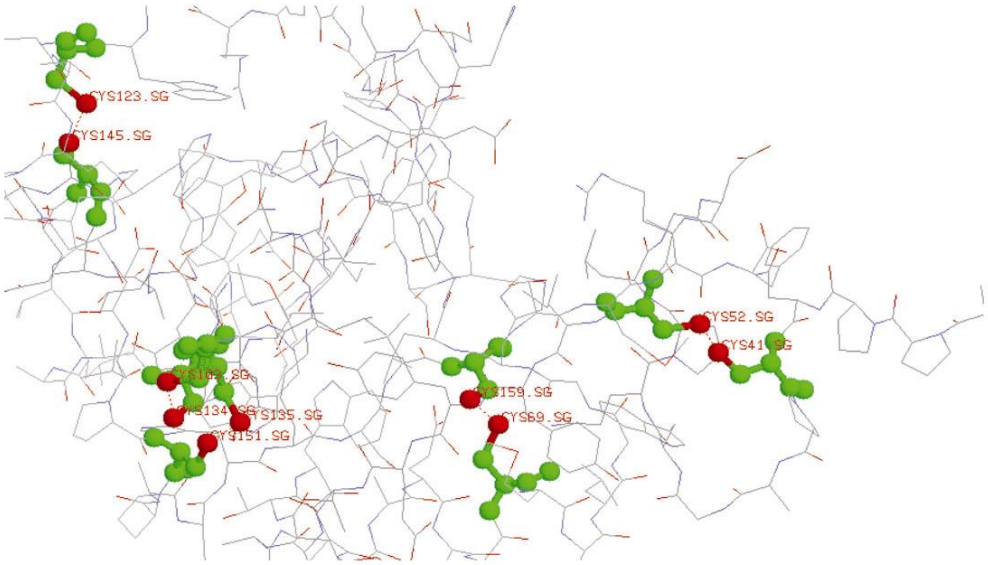

Figure 4. RasMol (wireframe diagram) representation of the homology modelled 3D structure of antifreeze protein P05140 (using PDB template 2AFP_A). The 10 cysteines are shown as ball and stick models (green). The sulphur atoms present in cysteines and the SS bonds (dotted lines) are shown in red colour. One unpaired cysteine is not shown.

P04002, P09031, Q8JI37 (Flounder fishes) and Q56TU0, P11920, P02732, P24856 (Cod fishes) have rich alanine content and mostly α-helices. AFPs P24028, P12101, P35751 (eelpout fishes), P12417 (Atlantic wolffish) Q01758 (rainbow smelt) P05140 (sea raven) and P07457(Ocean pout) have mixed secondary structure, i.e. α-helices β-strands and coils. The very high coil structural content of rainbow smelt fish (47·1%) and sea raven fish (50·4%) AFPs are due to the rich content of more flexible glycine and hydrophobic proline amino acids. Proline has a special property of creating kinks in polypetide chains and disrupting ordered secondary structure. The server SOSUI classifies the flounder fish AFPs P04002, P09031 and Atlantic cod fish AFP Q56TU0 as membrane protein and other AFPs as soluble proteins. SOSUI server has identified one transmembrane region in P04002 and Q56TU0 and two transmembrane regions in P09031. The transmembrane regions and their length are tabulated in table 5. The transmemebrane regions are rich in hydrophobic aminoacids and it is also well documented by Kyte and Dolittle mean hydrophobicity profile (figure 1) in which all the points are above the 0·0 line. The helix of Q56TU0 visualized using EMBOSS pepwheel is shown in figure 2. The tool CYS_REC recognizes the presence of 11 Cysteines in AFPs Q01758 and P05140 sequences and predicted five most probable SS bond pattern of pairs (as discussed in the primary structure analysis) in both of the proteins. The positions of five most probable SS bonds predicted by CYS_REC and the five SS bonds identi-

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

RamPage Percentage of residues in favoured region	CE RMSD (Å)	ProQ		Quality of the model
		LG score	Maxsub	
98	<2	>1·5	>0·1	Fairly good model
		>2·5	>0·5	Very good model
		>4	>0·8	Extremely good model

fied using Rasmol in the AFPs Q01758 and P05140 are shown in table 6. The three-dimensional structures of AFPs Q01758 and P05140 were modelled using various PDB templates (table 7) selected from the hits obtained through the BLASTP analysis and the modelled structures were evaluated. According to evaluation analysis, the Ramachandran plot and other parameters (table 8) were within the standard acceptable limits for the 3D structures modelled using the PDB template 2AFP_A for both of the (target) proteins. Criteria for a good 3D structure is given in table 9. The cysteines and the SS bonds identified using the three-dimensional structures of AFPs Q01758 and P05140 are shown in figures 3 and 4 respectively. In the case of AFP Q01758 the four SS bond positions Cys66-Cys135, Cys103-Cys145, Cys123-Cys134 and Cys151-Cys159 predicted by CYS_REC are not correlating with the SS bond positions Cys66-Cys159, Cys135-Cys151, Cys103-Cys134 and Cys123-Cys145 identified using Rasmol tool. We speculate that the SS bonds predicted from the primary structure (protein sequence) using CYS_REC tool might not be correct and the SS bonds identified from the three-dimensional structure (3D coordinates) using the Rasmol tool might be correct. The ten cysteines and five SS bonds present in the AFPs Q01758 and P05140 are shown in figures 3 and 4. The one unpaired cysteine in both the proteins is not shown in the figures.

4. Conclusions

Seventeen fish antifreeze proteins have been chosen mainly to study their physico-chemical properties, primary and secondary structures by using computational tools and servers. Primary structure analysis reveals that most of the AFPs under study are hydrophobic in nature and two of them contain disulphide linkages. Physico-chemical characterization studies give a good idea about the properties such as pI, EC, AI, GRAVY and Instability Index that are essential and vital in providing data about the proteins and their properties. Secondary structure analysis predicts that most of them contain only α-helices and remaining of them contain mixed structure. The presence of 11 Cys residues in rainbow smelt fish and sea raven fish indicates the presence of disulfide bridges which is also confirmed using CYS_REC and Rasmol tools.

References

1. Sivakumar K 2005 *Adv. BioTech.* **IV** 27
2. Sivakumar K, Balaji S and Gangaradhakrishnan *Bio-Chemistry: An Indian Journal* (in press)
3. Sivakumar K 2006 *Adv. BioTech.* **IV** 18
4. King-Hwa Ling, Shu-San Loo, Rozita Rosli, Mariana Nor Shamsudin, Rahmah Mohamed and Kiew-Lian Wan 2007 *Silico Biol.* **7** 0011
5. Chitta Suresh Kumar, Anuradha C M, Venkata Rao K and Venkateswara Swamy K 2005 *Int. J. Genomics Proteomics* **2** 1
6. Yuri F, Bogdanov, Sergei Y, Dadashev and Tatiana M Grishaeva 2003 *Silico Biol.* **3** 0015
7. Courtney E Garry and Robert F Garry 2004 *Theor. Biol. Med. Model.* **1** 10
8. Rachel E Bell and Nir Ben-Tal 2003 *Comp. Funct. Genom.* **4** 420
9. Raymond J A and DeVries A L 1977 *Proc. Natl. Acad. Sci. USA* **74** 2589
10. Peter L Davies and Choy L Hew 1990 *The FASEB J.* **4** 2460
11. Knight C A, DeVries A L and Oolman L D 1984 *Nature* (*London*) **308** 295
12. Garth L, Fetcher, Sally V Goddard and Yaling Wu 1999 *Chemtech.* **30** 17
13. Chao H, Davies P L and Carpenter J F 1996 *J. Exp. Biol. Sep.* **199** 2071
14. Parody-Morreale A, Murphy K P, Di Cera E, Fall R, DeVries A L and Gill S J 1988 *Nature* (*London*) **333** 782
15. Steffen P, Graether Carolyn M Slupsky Peter L, Davies and Brian D Sykes 2001 *Biophys. J.* **81** 1677
16. Pranav Dalal and Frank D Sonnichsen 2000 *J. Chem. Inf. Comput. Sci.* **40** 1276
17. Li Tong, Qingsong Lin, Raymond Wong. W K, Asma Ali, Daniel Lim, Wing L Sung, Choy L Hew and Danie S C Yang 2000 *Protein expression and purification* **18** 175

18. Wilkins S P, Bluma A J, Burkepilea D E, Rutland T J Wierzbicki A, Kelly M and Hamanna M T 2002 *CMLS, Cell. Mol. Life Sci.* **59** 2210
19. Boeckmann B, Bairoch A, Apweiler R, Blatter M-C, Estreicher A, Gasteiger E, Martin M J, Michoud K, O'Donovan C, Phan I, Pilbout S and Schneider M 200, *Nucl. Acids Res.* **31** 365
20. CLC bio., 2006. CLC free Workbench. http://www.clcbio.com/index.php?id=28, (27/10/2006)
21. Gill S C and Von Hippel P H 1989 *Anal. Biochem.* **182** 319
22. Bachmair A, Finley D and Varshavsky A 1986 *Science* **234** 179
23. Gonda D K, Bachmair A, Wunning I, Tobias J W, Lane W S and Varshavsky A 1989 *J. Biol. Chem.* **264** 16700
24. Tobias J W, Shrader T E, Rocap G and Varshavsky A 1991 *Science* **254** 1374
25. Ciechanover A and Schwartz A L 1989 *Trends Biochem. Sci.* **14** 483
26. Guruprasad K, Reddy B V B and Pandit M W 1990 *Prot. Eng.* **4** 155
27. Ikai A 1980 *J. Biochem.* **88** 1895
28. Kyte J and Doolittle R F 1982 *J. Mol. Biol.* **157** 105
29. Combet C, Blanchet C, Geourjon C and Deléage G 2000 *TIBS* **25** 291, 147
30. Eisenhaber F, Imperiale F, Argos P and Froemmel C 1996 *Proteins Struct. Funct. Design* **25** 157
31. Takatsugu Hirokawa, Seah Boon-Chieng and Shigeki Mitaku 1998 *Bioinform. Appl. Note* **14** 378
32. Ramachandran G N and Sasiskharan V 1968 *Adv. Prot. Chem.* **23** 283
33. CYS_REC. http://sun1.softberry.com/berry.phtml?topic=cys_rec&group=help &subgroup=propt. (27/10/2006)
34. Lambert C, Leonard N, De Bolle X and Depiereux E 2002 *Bioinformatics* **18** 1250
35. Lovell. S C, Davis I W, Arendall III W B, de Bakker P I W, Word J M, Prisant M G, Richardson J S and Richardson D C 2002 *Proteins: Structure, Function & Genetics* **50** 437
36. Cristobal S, Zemla A, Fischer D, Rychlewski L and Elofsson A 2001 *BMC Bioinformatics* **2** 5
37. Ilya N Shindyalov and Philip E Bourne 2001 *Nucl. Acids Res.* **29** 228