INT. J. BIOAUTOMATION, 2012, 16(4), 225-238

In silico Characterization of Plant and Microbial Antifreeze Proteins

Md. Musharaf Hossain[*], Abu Taher Mohammad Jannatul Mosnaz, Abdul Mohin Sajib, Jewel Halder, Raihan Mahmud, Khairul Alam Mohammad Nayeem, Nabangshu Shekhar Das, Muzahidul Islam Sarkar, Janmajoy Dey, Akhikun Nahar

Department of Genetic Engineering and Biotechnology
Shahjalal University of Science and Technology
Sylhet-3114, Bangladesh
E-mail: mmhossain27@gmail.com

[*]*Corresponding author*

Received: June 26, 2012 **Accepted: December 28, 2012**

Published: January 8, 2013

Abstract: *Antifreeze proteins (AFPs) are class of proteins that protect organisms from the damage caused by freezing through their ability to inhibit ice growth and effectively lower the temperature at which water freezes. In this study, a total of 25 antifreeze proteins were selected from four different sources (plant, bacteria and fungus) where they represent distinct physicochemical and structural features. Several Physico-chemical properties such as grand average hydropathy (GRAVY), aliphatic index (AI), extinction coefficient (EC), isolelectric point (pI), and instability index (II) were computed. S-S bridges and secondary structures were analyzed using CYS_REC and SOPMA programs respectively. The three dimensional structure of Antifreeze proteins is predicted by using three homology modelling server Geno3D, Swiss-model and CPHmodels. These models were 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.*

Keywords*: Antifreeze Proteins, Hydrophobicity, Homology modelling, In silico analysis, Isoelectric point.*

Introduction

Freezing is almost always lethal to cellular organisms as it deprives biological reactions of the aqueous medium they require, causes concentration of ions and other solutes in the plasma, denaturation of biomolecules, and ruptures cell membranes [1]. Different species vary enormously in their ability to withstand cold and freezing temperature where they possess some special features. Cold tolerance plants has some kind of proteins, these fall into a small number of groups, but they all share the property of being extremely hydrophilic. Some proteins contain repeat motif and some are late embryogenesis abundant (LEA) proteins. Other groups of proteins are encoded by a class of genes designated as cold responsive (COR) genes according to their pattern of expression. Many species of ectothermic animals, plants, and microbes living in cold environments produce antifreeze proteins/polypeptides (AFPs) to protect them from freezing damage [2-4] by possessing thermal hysteresis (TH) and recrystallization inhibition (RI) activity. It is generally accepted that AFPs function through adsorption of their flat ice-binding surfaces onto particular planes of ice crystals and prevent or inhibit further ice growth [5]. However, AFP-producing plants and bacteria reported to date show substantially lower thermal hysteresis activity than do animals but they prefer RI

activity of AFPs to control the size of ice crystals. Various structurally distinct AFPs have evolved independently [6] and to date a total of 5 types of structurally distinct AFPs like antifreeze glycoprotein's, type I, type II, type III, and type IV were identified. Because of their recrystallization inhibition property, AFPs are highly useful in the preservation technique that's why it has potential applications in cryosurgery of tumors, transplantation, transfusion [7] and as a component of ice-cream to prevent the formation of hard and large ice crystals [8]. To date several types of antifreeze proteins were purified and analyzed from different sources to resolve the protein-ice interaction [9], evolution of AFPs [10, 11], structure function correlation [12], molecular dynamics and modeling studies [13].

Nowadays several computational approaches, software, algorithms and online servers are available for genomic, proteomic and evolutionary analysis to accelerate experimental outcomes as well as widening scientific thoughts. Computational tools provide researchers a cost effective way to understand physicochemical and the structural aspects of a protein for the successful design of many biological experiments with in a short range of time and these methods are not amenable to high throughput techniques. Physicochemical characterization studies give more information about the properties such as M.wt, pI, AI, GRAVY and Instability Index. These properties are essential and vital for the characterization of proteins and their properties [14]. Numerous structure and function studies of AFPs have been reported experimentally from time to time while computational study of AFPs are much more limited. Hence to describe its structural features and to understand molecular function, the model structures for these proteins were constructed. In this study, we will focus on the *in silico* characterization and homology modeling of AFPs from different sources

Materials and methods

Sequence retrieval from Swiss-Prot

Sequences of antifreeze protein were retrieved from Swiss-Prot, a public domain protein database [15]. A total of 25 sequences (15 bacterial, 4 fungal and 6 sequences from plant) were retrieved from Swiss-Prot having protein sequences from different spectrum at different number by random selection. Table 1 shows the protein sequences considered in this study.

Tools and servers used for in silico study

The amino acid compositions of all retrieved protein sequences of 3 different sources were determined (Table 2 and Table 3). The physico-chemical properties such as, theoretical pI, molecular weight, total number of positive and negative residues, EC [16], half-life [17], instability index [18], aliphatic index [19] and grand average hydrophathy (GRAVY) [20] were computed using the Expasy's ProtParam prediction server (http://us.expasy.org/tools/protparam.html) [21]. The SOSUI server was used for the identification of transmembrane regions of a protein. The predicted transmembrane helices were visualized and analyzed using Helical Wheel Plots. Disulphide bonds are very essential in determining the functional linkage and the stability of a particular protein. The presence of S-S bond and their bonding pairs were predicted by CYS_REC (http://linux1.softberry.com/berry phtml?topic) and "What If" server (identify SS bonds from 3D structure of a protein). The secondary structural features of antifreeze proteins were calculated by employing SOPMA [22].

Table 1. Antifreeze protein sequences considered for the study

Accession number	Sequence description	Organism
A3WTB9	Antifreeze-like protein	*Nitrobacter sp*. Nb-311A
DOAWG3	Antifreeze protein	*Brucella abortus* NCTC 8038
D9XVH6	Type I antifreeze protein	*Streptomyces griseoflavus* Tu4000
Q68VA9	Antifreeze protein	*Pseudomonas putida*
D9VES8	Antifreeze protein	*Streptomyces sp*. AA4
A1B688	Antifreeze protein, type I	*Paracoccus denitrificans* (strain Pd 1222)
A9KKC7	Antifreeze protein type I	*C. phytofermentans* (strain ATCC 700394)
Q1IVR1	Type I antifreeze protein	*Acidobacteria bacterium* (strain Ellin345)
B5YFN9	Type I antifreeze protein	*T. yellowstonii* (strain ATCC 51303)
Q8P6Q0	Antifreeze glycopeptide	*Xanthomonas campestris pv. campestris*
Q82VH2	Type I antifreeze protein	*Nitrosomonas europaea*
Q6N3V0	Type I AFP	*Rhodopseudomonas palustris*
A4YMB9	Type I antifreeze protein	*Bradyrhizobium sp.* (strain ORS278)
D8IRL6	Type I antifreeze protein	*Herbaspirillum seropedicae* (strain SmR1)
B8I1H8	Antifreeze protein type I	*C. cellulolyticum* (strain ATCC 35319)
Q9P3M2	AFGP polyprotein	*Neurospora crassa*
C7F6X3	Antifreeze protein	*Leucosporidium sp.* AY30
Q76CE2	Antifreeze protein	*Typhula ishikariensis*
DOEKL2	Antifreeze protein Afp1	*Leucosporidium antarcticum*
Q9AXR8	Class II endochitinase AFP	*Secale cereale* (Rye)
Q9AXR9	Class I endochitinase AFP	*Secale cereale* (Rye)
Q9S9D9	AFA3	*Nicotiana tabacum* (Common tobacco)
Q6H6G5	AFGP related protein	Oryza sativa subsp. japonica
Q53LT2	AFGP polyprotein	Oryza sativa subsp. japonica
Q6UAH5	Antifreeze protein	Populus suaveolens

Table 2. Amino acid composition of plant and fungal antifreeze proteins (in %)

	Plant						Fungi			
AMINO ACIDS	Q9AXR8	Q9AXR9	Q9S9D9	Q6H6G5	Q53LT2	Q6UAH5	Q9P3M2	C7F6X3	Q76CE2	D0EKL2
Ala	13.5%	11.0%	60.5%	17.9%	26.6%	4.6%	18.50%	13.00%	13.60%	14.10%
Arg	6.3%	4.7%	2.6%	8.9%	2.3%	9.3%	3.50%	2.30%	0.80%	4.00%
Asn	5.6%	4.4%	5.3%	2.4%	3.1%	0.7%	4.50%	4.20%	2.10%	1.10%
Asp	5.6%	4.7%	7.9%	9.8%	6.0%	4.0%	4.20%	5.00%	2.50%	6.20%
Cys	2.0%	5.3%	0.0%	2.4%	1.7%	2.0%	0.30%	0.00%	0.00%	1.70%
Gln	3.2%	4.4%	0.0%	0.8%	2.3%	9.9%	2.80%	4.60%	2.90%	2.30%
Glu	2.0%	1.9%	0.0%	8.1%	5.1%	5.3%	5.90%	3.10%	1.60%	1.70%
Gly	11.9%	12.6%	0.0%	11.4%	6.6%	7.9%	11.90%	11.10%	13.20%	8.50%
His	0.8%	1.6%	0.0%	0.0%	0.9%	2.6%	0.00%	0.40%	0.00%	1.70%
Ile	3.6%	3.5%	0.0%	1.6%	3.1%	5.3%	2.40%	4.20%	7.40%	4.00%
Leu	5.6%	4.7%	5.3%	9.8%	8.0%	7.3%	4.90%	10.70%	9.10%	13.00%
Lys	2.0%	2.5%	2.6%	1.6%	3.4%	6.6%	7.00%	2.30%	3.30%	2.30%
Met	2.8%	1.3%	2.6%	1.6%	1.4%	2.6%	2.10%	0.40%	0.80%	0.60%
Phe	4.8%	4.7%	0.0%	0.8%	0.9%	7.9%	2.40%	4.20%	3.30%	5.60%
Pro	4.0%	6.6%	0.0%	11.4%	6.3%	2.6%	7.00%	4.20%	3.30%	4.50%
Ser	6.0%	8.2%	2.6%	3.3%	6.6%	7.9%	5.90%	8.40%	11.10%	12.40%
Thr	8.7%	6.0%	10.5%	0.8%	6.6%	5.3%	9.80%	8.80%	13.60%	7.90%
Trp	1.6%	1.9%	0.0%	2.4%	0.3%	2.0%	0.00%	1.10%	1.20%	0.60
Tyr	4.8%	4.7%	0.0%	0.8%	3.4%	1.3%	1.70%	2.70%	2.10%	1.70%
Val	5.6%	5.3%	0.0%	4.1%	5.4%	4.6%	4.90%	9.20%	8.20%	6.20%

Table 3. Amino acid composition of bacterial antifreeze proteins (in %)

AMINO ACIDS	A3WTB9	D8IRL6	D0AWG3	D9XVH6	Q68VA9	D9VES8	A1B688	A9KKC7	Q1IVR1	B5YFN9	Q8P6Q0	Q82VH2	Q6N3V0	A4YMB9	B8I1H8
Ala	13.70%	16.40%	14.60%	9.20%	17.30%	10.40%	9.00%	5.80%	8.10%	1.20%	26.10%	17.70%	11.80%	16.60%	7.90%
Arg	4.90%	7.20%	4.40%	2.80%	1.50%	4.20%	5.60%	2.50%	3.00%	2.40%	5.60%	8.40%	9.40%	7.80%	3.20%
Asn	2.30%	1.40%	3.70%	1.80%	7.20%	3.40%	3.50%	8.10%	0.00%	7.30%	1.60%	0.80%	0.50%	0.50%	5.30%
Asp	4.90%	4.70%	8.50%	2.80%	8.00%	4.70%	4.80%	4.20%	4.00%	1.20%	5.90%	5.10%	2.80%	5.70%	6.10%
Cys	1.10%	0.00%	0.20%	3.70%	0.00%	0.00%	0.00%	4.60%	4.00%	4.90%	0.00%	0.00%	0.00%	0.00%	3.70%
Gln	2.60%	8.60%	6.30%	3.70%	3.40%	10.70%	4.80%	2.80%	3.00%	3.70%	7.20%	8.10%	7.10%	3.10%	2.40%
Glu	6.60%	5.00%	10.00%	3.70%	4.90%	5.50%	6.10%	6.00%	12.10%	8.50%	2.90%	6.70%	11.80%	4.70%	4.70%
Gly	8.60%	6.10%	7.10%	11.00%	11.20%	9.10%	10.90%	9.50%	6.10%	7.30%	3.30%	5.10%	19.30%	8.80%	8.70%
His	2.90%	1.10%	2.20%	0.00%	0.80%	0.00%	1.90%	0.90%	0.00%	2.40%	1.90%	1.10%	1.90%	4.10%	0.80%
Ile	8.90%	3.10%	2.90%	0.90%	6.30%	5.20%	6.10%	7.40%	1.00%	6.10%	2.40%	6.20%	2.80%	4.10%	9.50%
Leu	8.60%	10.80%	4.50%	2.80%	6.80%	7.30%	6.90%	6.50%	3.00%	3.70%	5.60%	8.70%	3.80%	11.90%	7.10%
Lys	5.10%	3.30%	1.10%	4.60%	0.60%	3.90%	2.90%	7.90%	15.20%	12.20%	1.00%	3.10%	7.10%	2.10%	7.90%
Met	1.40%	1.70%	1.50%	0.90%	0.20%	3.40%	3.70%	1.80%	1.00%	2.40%	2.10%	2.00%	0.90%	2.60%	1.60%
Phe	3.40%	2.20%	3.60%	3.70%	3.00%	4.20%	4.80%	4.20%	3.00%	2.40%	1.60%	1.70%	0.50%	3.10%	5.80%
Pro	5.40%	4.20%	9.00%	3.70%	1.70%	8.10%	8.20%	5.30%	7.10%	3.70%	11.50%	4.20%	15.10%	7.30%	4.70%
Ser	6.00%	5.80%	6.80%	31.20%	8.20%	3.90%	3.50%	6.50%	15.20%	13.40%	9.40%	4.50%	0.90%	4.10%	4.70%
Thr	4.90%	5.30%	4.50%	7.30%	8.90%	4.70%	6.40%	5.30%	5.10%	4.90%	5.30%	3.90%	0.90%	4.70%	5.80%
Trp	1.10%	0.80%	0.70%	0.00%	0.80%	1.80%	1.90%	0.90%	1.00%	1.20%	1.30%	0.60%	0.50%	0.00%	0.50%
Tyr	3.40%	1.10%	3.00%	2.80%	1.70%	2.90%	2.70%	4.20%	4.00%	4.90%	0.30%	1.40%	0.90%	3.60%	3.20%
Val	4.30%	11.10%	5.20%	3.70%	7.40%	6.80%	6.40%	5.80%	4.00%	6.10%	5.30%	10.70%	1.90%	5.20%	6.60%

Comparative modeling and evaluation

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 [23]. The modeling of the protein was done to deduce the three dimensional structure of the protein. Structure prediction of a plant antifreeze protein (Q9AXR9) has been based on the availability of existing solved template structure, which has sequence homology with the target sequences. Homology modeling of this protein was done by using a template structure (2DKV_A) from PDB (http://www.pdb.org/pdb/home/home.do) through BLASTP search (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Three homology modeling programs Geno3D [24], Swiss-model [25], CPHmodels [26] were used for the modeling of the target antifreeze protein. The modeled 3D structure was evaluated using the online server Rampage, ProQ (Protein quality server) and ProSA. The structure validation of antifreeze proteins was performed by online PROCHECK [27] and "What If" [28] and visualized with Pymol.

Results and discussion

Table 1 shows a total of 25 AFPs considered in this study. These protein sequences were retrieved from Swissprot database and different physiological features (molecular weight (M. wt.), isoelectric point (pI), number of positive (+R) and negative residues (-R), extinction coefficient (EC), instability index (II), aliphatic index (AI) and grand average hydrophathy (GRAVY)) were computed using EsPasy's protparam was represented in Table 4. The total number of amino acid residues ranged from 38 to 990 with variable molecular weights. The result of primary structure analysis infers that, proteins from fungi and plant are mostly hydrophobic and AFPs from bacterial species are mostly hydrophilic. The assumed series of hydrophobicity is Fungi > Plant > Bacteria (according to analyzed protein sequences) and their hydrophobic nature is due to the presence of high non-polar residues. Sivakumar et al. [29] also suggests that due to the high content of non-polar residues most of the AFPs are hydrophobic in nature. The presence of 20 (4.6%) Cys residues in A9KKC7 (*Clostridium phytofermentans*), 14 (3.7%) Cys in B8I1H8 (*Clostridium cellulolyticum*), 17 (12.6%) Cys in X9AXR9 (*Secale cereale*) indicate the presence of disulphide bonds in corresponding Antifreeze protein. Moreover, the primary structure also suggests that the AFP Q9S9D9 have no aromatic residues (Phe, Trp and Tyr). Isoelectric point (pI) is the pH at which the net charge of the protein is zero and at pI proteins are compact and stable. The computed pI value of AFPs ranged from 3.67 (Q68VA9) to 10.02 (Q6UAH5) where the former one is basic and the later one is acidic in character. Most plant antifreeze proteins have basic character (according to retrieved protein sequences) with pI value in average of 6.555. On the other side, Bacteria and fungi contain AFPs with acidic nature mostly. For the purification of a particular protein by isoelectric focusing methods, the pI value of this protein will be useful for developing buffer system.

Extinction Coefficient (EC) of AFPs were calculated by EsPasy protparam at 280 nm wavelength is ranging from 1490 to 83325 $M^{-1}\cdot cm^{-1}$ with respect to the concentration of Cys, Trp and Tyr and it helps in the quantitative study of protein-protein and protein-ligand interactions in solution. The high EC value of DOAWG3, Q9AXR9, D9VES8, A1B688 and A9KKC7 indicates presence of high concentration of Cys, Trp and Tyr. EsPasy protparam computes no EC value for Q9S9D9 due to the absence of Cys, Trp and Tyr. This indicates that these AFPs can not be analyzed using UV spectral methods. 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's whose instability index < 40 is predicted as stable; a value above 40 predicts that the protein may be unstable [18]. 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 34.13 (D9XVH6) to 106.71 (Q82VH2) among retrieved sequences. The lower thermal stability of D9XVH6, Q1IVR1 and Q6N3V0 is indicative of a more flexible structure when compared to other AFPs. The very high aliphatic index of Q92006, Q1AMQ2, Q1AMQ6 and Q1AMQ8 infers that these AFPs may be stable for a wide range of temperature. The instability index of AFPs are ranging from 1.31 to 67.97 and suggesting bacterial antifreeze proteins as mostly unstable protein among others from different organisms.

Table 4. Physicochemical parameters of different AFPs computed using Expasy's ProtParam

Accession number	Length	M. wt.	pI	(-) R	(+) R	EC	II	AI	GRAVY
Q9AXR8	252	26835	8.25	19	21	40130	32.11	65.2	-0.17
D8IRL6	360	38443	9.15	35	38	22460	29.9	102.78	0.027
Q9AXR9	318	33636.6	7.82	21	23	56350	41.39	58.4	-0.194
Q9S9D9	38	3359.6	4.43	3	2	NIL	1.31	81.05	0.563
A3WTB9	350	37787.2	6.08	40	35	40130	32.7	94.11	-0.036
DOAWG3	990	106361.3	4.12	183	55	83325	56.52	58.74	-0.695
D9XVH6	109	10754.4	7.62	7	8	4720	67.97	34.13	-0.483
Q68VA9	473	47316.1	3.67	61	10	33920	12.34	89.92	0.083
D9VES8	384	42103.8	4.88	39	31	54890	47.68	78.8	-0.307
A1B688	376	41168.8	5.38	41	32	53400	40.49	78.38	-0.221
A9KKC7	433	47427.1	7.33	44	45	50070	42.66	76.56	-0.268
Q1IVR1	99	10879.1	8.31	16	18	11710	45.38	35.56	-1.074
B5YFN9	82	9250.5	8.95	8	12	11710	57.58	56.95	-0.765
Q8P6Q0	628	63632.7	5.32	55	41	46980	64.48	72.4	-0.193
Q82VH2	356	38415.1	6.49	42	41	18450	41.92	106.71	0.023
Q6N3V0	212	22070.5	9.32	31	35	8480	36.47	43.02	-1.293
A4YMB9	193	20672.6	6.71	20	19	10430	36.85	94.25	-0.025
B8I1H8	380	41517	7.39	41	42	29755	28.03	91.63	0.044
Q9P3M2	286	28688.9	7.71	29	30	7450	30.68	61.36	-0.388
C7F6X3	261	26807.1	4.43	21	12	26930	32.64	97.97	0.262
Q76CE2	243	23989.3	6.15	10	10	23950	20.14	101.65	0.566
DOEKL2	177	18243.7	5.45	14	11	10095	39.94	98.25	0.394
Q6H6G5	123	12827.2	4.41	22	13	18115	67.16	74.07	-0.446
Q53LT2	350	34782.8	4.40	39	20	23755	38.13	85.77	0.129
Q6UAH5	151	17548.1	10.02	14	24	19605	55.42	67.15	-0.550

Note: molecular weight (**M. wt.**), isoelectric point (**pI**), number of positive (+**R**) and negative residues (**-R**), extinction coefficient (**EC**), instability index (**II**), aliphatic index (**AI**) and grand average hydrophathy (**GRAVY**).

Table 5. Transmembrane region along with their length and type identified by SOSUI server

Accession number	Transmembrane region	Type	Length
Q76CE2 (FUNGUS)	SASSLLAVIALAISSVSAAGPSA	PRIMARY	23
	LGTAGNYVILASTGVSTVPQSVI	SECONDARY	23
	TGFSLILSGTGTFSTSSQVTGQL	SECONDARY	23
Q9AXR9 (PLANT)	VVVVAMLAAAFAVSAHAEQCG	PRIMARY	21
	AKGFYNYGAFIAAANSFSAFATT	SECONDARY	23
Q53LT2 (PLANT)	AAAPITLLVLSLLLLAVAAATAA	PRIMARY	23
	AAGANAASNIAAGAAAGMAADAA	SECONDARY	23
Q6H6G5 (PLANT)	APAGAALALAAAVCFLLMAPAPA	PRIMARY	23
Q9AXR8 (PLANT)	AALAALLLAVAVGGAAAQSVGSV	PRIMARY	23
D8IRL6 (BACTERIA)	PLLAAAALLLLLGAAFLAWQRLH	PRIMARY	23
D0AWG3 (BACTERIA)	GLILASVAGVAVLLGGIGYHFLG	PRIMARY	23
Q82VH2 (BACTERIA)	KPVLTYLLAALVIITIIVAWRVL	PRIMARY	23

Table 6. Secondary structure features calculated by SOPMA

Source	Accession number	Secondary structure features			
		Alpha helix	Extended	Beta turn	Random coil
BACTERIA	A3WTB9	42.00%	17.43%	6.29%	34.29%
	D0AWG3	29.60%	8.89%	5.35%	56.16%
	D9XVH6	17.43%	10.09%	4.59%	67.89%
	Q68VA9	39.53%	20.51%	10.36%	29.60%
	D9VES8	35.42%	18.23%	7.55%	38.80%
	A1B688	37.50%	19.15%	7.18%	36.17%
	A9KKC7	29.10%	16.40%	6.93%	47.58%
	Q1IVR1	22.22%	9.09%	6.06%	62.63%
	D8IRL6	51.67%	15.56%	5.83%	26.94%
	B5YFN9	23.17%	21.95%	4.88%	50.00%
	Q8P6Q0	37.90%	8.28%	4.14%	49.68%
	Q82VH2	51.40%	16.29%	5.62%	26.69%
	Q6N3V0	31.13%	3.30%	1.42%	64.15%
	A4YMB9	33.16%	20.21%	7.77%	38.86%
	B8I1H8	31.84%	17.11%	7.37%	43.68%
FUNGI	Q9P3M2	35.31%	6.64%	3.85%	54.20%
	C7F6X3	17.62%	30.65%	10.73%	41.00%
	Q76CE2	17.70%	28.40%	9.88%	44.03%
	D0EKL2	41.24%	15.25%	9.04%	34.46%
PLANT	Q9AXR8	31.75%	15.48%	5.56%	47.22%
	Q9AXR9	23.90%	12.26%	5.03%	58.81%
	Q9S9D9	92.11%	0.00%	0.00%	7.89%
	Q6H6G5	27.64%	6.50%	4.07%	61.79%
	Q53LT2	61.71%	3.71%	5.14%	29.43%
	Q6UAH5	24.50%	25.17%	10.60%	39.74%

Note: All other secondary structure features such as 310 helix, Pi helix, Ambigous states, Bend region and Beta bridge were not found.

The highest instability index value was obtained from D9XVH6 (67.97) which is followed by Q6H6G5 (67.16), Q8P6Q0 (64.48), B5YFN9 (57.58), DOAWG3 (56.52) and so on. Contrarily, the lowest instability index value was obtained from AFP Q9S9D9 (1.31) of tobacco plant and Q68VA9 (12.34) of *Pseudomonas putida*. The Grand Average Hydropathy (GRAVY) value of all AFPs are ranging from -1.293 (Q6N3V0) to 0.566 (Q76CE2) and infers that the AFPs from bacteria are mostly hydrophilic due to low GRAVY index. The very low GRAVY index of AFPs Q6N3V0, Q1IVR1 and B5YFN9 indicates the possibility of better interaction with water.

Protein classification and transmembrane helices identification

Functional characterization of antifreeze protein was also performed including transmembrane (TM) region identification, prediction of disulphide bonding pairs etc. along with physiochemical characterization. The SOSUI server performed the identification transmembrane helices with their corresponding length and differentiates membrane proteins from stable proteins. The server SOSUI classifies Q9AXR9, Q53LT2, Q6H6G5, Q9AXR8, D8IRL6, DOAWG3, Q82VH2 and Q76CE2 as membrane protein and others as soluble proteins.

All these antifreeze membrane proteins contain distinct number TM helices such as Q76CE2 (*Typhula ishikariensis*) contains 3 TM helices, Q53LT2 (*Oryza sativa subsp. japonica*) and Q9AXR9 (Rye) contains 2 TM helices and rest of them all contains 1 TM region of each (Table 5). Transmembrane helices of these proteins were also identified by using TMHMM and TMpred (Transmembrane prediction) server. Hydrophobicity of these AFPs was

computed based on Kyte Dolittle hydrophobicity index by ProtScale (http://expasy.org/tools/protscale.html).

Table 7. 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"
Q9AXR9	**Cys 23- Cys 35**	**Cys 23- Cys 38**
	Cys 32- Cys 62	Cys 32- Cys 44
	Cys 37- Cys 51	Cys 37- Cys 51
	Cys 38- Cys 44	Cys 55- Cys 59
	Cys 55- Cys 59	Cys 98- Cys 160
	Cys 98- Cys 160	Cys 172- Cys 180
	Cys 172- Cys 180	Cys 279- Cys 311
	Cys 279- Cys 311	

Secondary structures analysis

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 [22]. This information gives an idea whether a given amino acid lies in a helix, strand or coil. The secondary structure were predicted by using default parameters (Windows width: 17, similarity threshold: 8, and number of states: 4). Table 6 represents the calculated secondary structure features of AFPs analyzed. This result infers 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 (Fig. 1). In case of plant AFPs, alpha helix is the dominating secondary structure feature. 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 were identified from 3D structure by "What If" in the AFP Q9AXR9 are shown in Table 7.

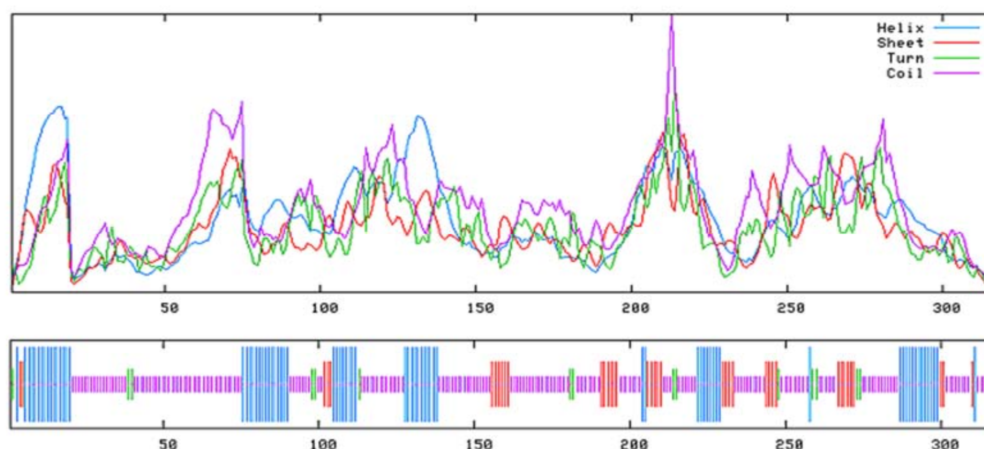

Fig. 1 Secondary structural features of antifreeze protein (Q9AXR9) predicted by SOPMA using default parameters

Homology modelling 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 modelling servers at different level of complexity.

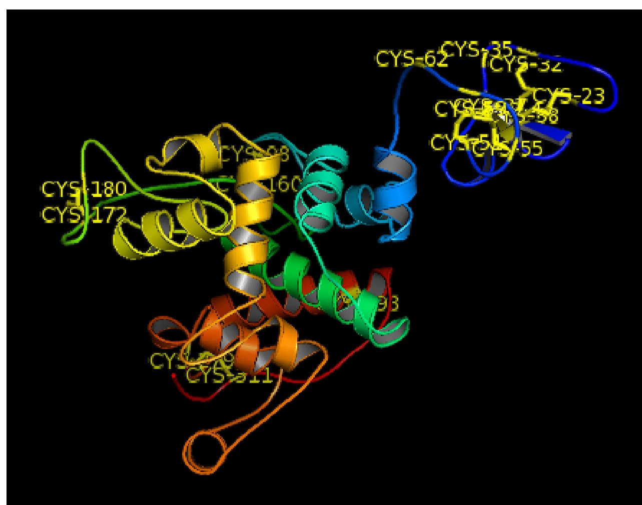

Fig. 2 Homology modeled 3D structure
of plant antifreeze protein Q9AXR9 (Rye)

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 [30]. The modeling of 3D structure of protein was performed by three homology modeling program Geno3D, Swiss model (Fig. 2) and CPHmodels. In this study, Q9AXR9 (Rye) is considered for homology modeling based on PDB template selected from the hits obtained through the $BLAST_P$ search. 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 [27]. 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 form well refined high resolution structure.

Accession number

Table 8. Ramachandran plot calculation, comparative analysis with PROCHECK program and validation parameters of homology modeled structure computed by ProQ and "What If" server

Accession number	Template	Server	Rampage analysis			ProQ result		
			RFR	**RAR**	**ROR**	**RMS Z score**	**LG score**	**Maxsub**
Q9AXR92	DKV_A	Swiss-model	95.6%	3.7%	0.7%	1.101	3.504	0.354
		Geno3D	83.7%	12.9%	3.4%	0.474	3.102	0.295
		CPHmodels	90.2%	6.4%	3.4%	0.916	3.721	0.352

Note: ROR (residues in favored region), RAR (residues in allowed region), ROR (residues in outlier region).

The result revealed that, the protein Q9AXR9 modeled by Swiss model homology modelling server has average maximum residues in favored region which are about 95.6% respectively. A comparison of the results obtained from three different modelling server in Table 8 shows that the models generated by Swiss model was more acceptable in comparison with others. The modeled structure of antifreeze protein was 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.

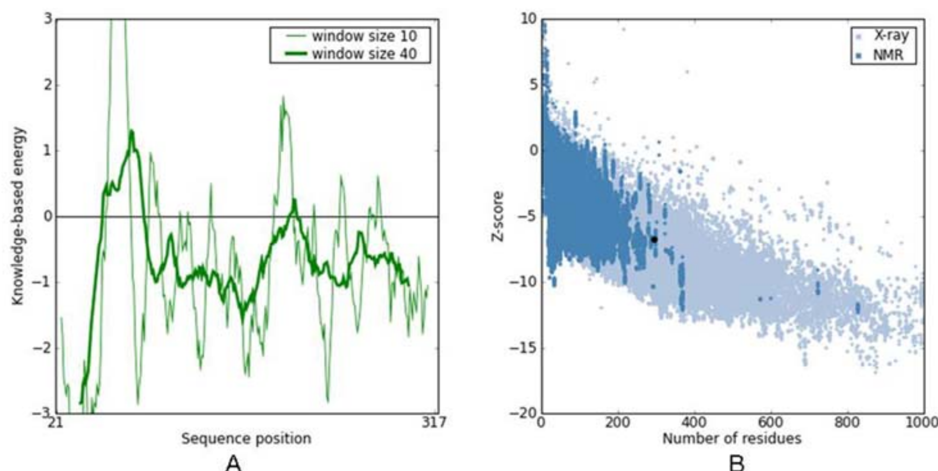

Fig. 3 ProSA-web service analysis of AFP Q9AXR9

ProQ is a neural network based predictor that based on a number of structural features predicts the quality of a protein model and optimized to find native structures whereas RMS Z score for angles of models are determined using "What If". Two quality measures, LG score and MaxSub of three models from each modelling server are predicted by ProQ and enlisted with RMS Z score and PDB templates in Table 8. 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 Plant (Q9AXR9) AFP was shown in Fig. 2. S-S bonding pairs predicted by CYS_REC are 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 Fig. 3(A) displays an energy plot of Q9AXR9. In general, positive values correspond to problematic or erroneous parts of a model. The energy plot shows the local model quality by plotting energies as a function of amino acid sequence position. Fig. 3(B) the Z-score of Q9AXR9 is also well within the range of scores typically found for proteins of similar size indicating a highly reliable structure.

References

1. Deng C., C. H. C. Cheng, H. Yea, X. He, L. Chen (2010). Evolution of an Antifreeze Protein by Neofunctionalization under Escape from Adaptive Conflict, Protocol of National Academy of Science, 107, 21593-21598.
2. Ewart K. V., Q. Lin, C. L. Hew (1999). Structure, Function and Evolution of Antifreeze Proteins, Cellular and Molecular Life Science, 55, 271-283.

3. Marshall C. B., G. L. Fletcher, P. L. Davies (2004). Hyperactive Antifreeze Protein in a Fish, Nature, 429, 153.
4. Graham L. A., P. L. Davies (2005). Glycine-rich Antifreeze Proteins from Snow Fleas, Science, 310, 461.
5. Fletcher G. L., C. L. Hew, P. L. Davies (2001). Antifreeze Proteins of Teleost Fishes, Annual Review of Physiology, 63, 359-390.
6. Davies P. L., B. D. Sykes (1997). Antifreeze Proteins, Current Opinion in Structural Biology, 7, 828-834.
7. Fletcher G. L., S. V. Goddard, Y. L. Wu (1999). Antifreeze Proteins and Their Genes: From Basic Research to Business Opportunity, Chemical Technology, 29, 17-28.
8. 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.
9. Jorov A., B. S. Zhorov, D. S. Yang (2004). Theoretical Study of Interaction of Winter Flounder Antifreeze Protein with Ice, Protein Science, 13, 1524-1537.
10. Sandve S. R., H. Rudi, T. Asp, O. A. Rognli (2008). Tracking the Evolution of a Cold Stress Associated Gene Family in Cold Tolerant Grasses, BMC Evolutionary Biology, 8, 245.
11. Davies P. L., B. D. Sykes (1997). Antifreeze Proteins, Current Opinion in Structural Biology, 7, 828-834.
12. Graether S. P., B. D. Sykes (2004). Cold Survival in Freeze-intolerant Insects: The Structure and Function of Beta-helical Antifreeze Proteins, European Journal of Biochemistry, 271, 3285-3296.
13. Lin F. H., L. A. Graham, R. L. Campbell, P. L. Davies (2007). Structural Modeling of Snow Flea Antifreeze Protein, Biophysical Journal, 92, 1717-1723.
14. Pradeep N. V., Anupama, K. G. Vidyashree, P. Lakshmi (2012). *In silico* Characterization of Industrial Important Cellulases using Computational Tools, Advances in Life Science and Technology, 4, 8-14.
15. Boeckmann B., A. Bairoch, R. Apweiler, M. C. Blatter, A. Estreicher, E. Gasteiger, M. J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout, M. Schneider (2003). The SWISS-PROT Protein Knowledgebase and Its Supplement TrEMBL in 2003, Nucleic Acids Research, 31, 365-370.
16. Gill S. C., P. H. Von Hippel (1989). Extinction Coefficient, Anal Biochemistry, 182, 319-328.
17. Tobias J. W., T. E. Shrader, G. Rocap, A. Varshavsky (1991). The N-end Rule in Bacteria, Science, 254(5036), 1374-1377.
18. Guruprasad K., B. V. P. Reddy, M. W. Pandit (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, Protein Engineering, 4, 155-164.
19. Ikai A. J. (1980). Thermo Stability and Aliphatic Index of Globular Proteins, Journal of Biochemistry, 88, 1895-1898.
20. Kyte J., R. F. Doolottle (1982). A Simple Method for Displaying the Hydropathic Character of a Protein, Journal Molecular Biology, 157, 105-132.
21. 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.
22. Geourjon C., G. Deléage (1995). SOPMA: Significant Improvements in Protein Secondary Structure Prediction by Consensus Prediction from Multiple Alignments, Computational and Applied Bioscience, 11, 681-684.
23. Schwede T., J. Kopp, N. Guex, M. C. Peitsch (2003). SWISS-MODEL: An Automated Protein Homology-modeling Server, Nucleic Acids Research, 13, 3381-3385.

24. Combet C., M. Jambon, G. Deleage, C. Geourjon (2002). Geno3D: Automatic Comparative Molecular Modelling of Protein, Bioinformatics, 18, 213-214.
25. Arnold K., L. Bordoli, J. Kopp, T. Schwede (2006). The SWISS-MODEL Workspace: A Web-based Environment for Protein Structure Homology Modeling, Bioinformatics, 22, 195-201.
26. Nielsen M., C. Lundegaard, O. Lund, T. N. Petersen (2010). CPHmodels-3.0-remote Homology Modeling using Structure-guided Sequence Profiles, Nucleic Acids Research, 10, 1093.
27. Laskowski R. A., J. A. Rullmannn, M. W. MacArthur, R. Kaptein, J. M. Thornton (1996). AQUA and PROCHECK-NMR: Programs for Checking the Quality of Protein Structures Solved by NMR, Journal of Biomolecular NMR, 8, 477-486.
28. Vriend G. (1990). WHAT IF: A Molecular Modeling and Drug Design Program, Journal of Molecular Graphics, 8, 52-56.
29. Sivakumar K., S. Balaji, Gangaradhakrishnan (2007). *In silico* Characterization of Antifreeze Proteins using Computationaltools and Servers, Journal of Chemical Science, 119(5), 571-579.
30. Chothia C., A. M. Lesk (1986). The Relation between the Divergence of Sequence and Structure in Proteins, EMBO Journal, 5(4), 823-826.

Md. Musharaf Hossain
E-mail: mmhossain27@gmail.com

Md. Musharaf Hossain graduated in genetic engineering and biotechnology from Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh in 2008 with distinction. Now, he is a working as a research assistant under an USDA project entitled “Improvement of Salt and Submergence Tolerant Rice to Genetic Engineering Approach to Bring Food Security with Environmental Safety in Bangladesh”. He has interest over many bioinformatics data analysis and structure prediction tools for drug designing, phylogenetic analysis and protein structure prediction.

Abu Taher Mohammad Jannatul Mosnaz
E-mail: mosnaz@gmail.com

A. T. M. J. Mosnaz completed his post graduation in genetic engineering and biotechnology from Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh in 2009. Now, he is working in an USDA project entitled “Improvement of Salt and Submergence Tolerant Rice to Genetic Engineering Approach to Bring Food Security with Environmental Safety in Bangladesh”. He has interest over plant tissue culture, molecular biology of viruses, bioinformatics data analysis and drug designing.

Abdul Mohin Sajib
E-mail: mohinick@yahoo.com

Abdul Mohin Sajib, graduated from Shahjalal University of Science & Technology, Sylhet under the department of Genetic Engineering & Biotechnology and has a good conceptual grasp on Microbiology, Plant and Animal Biotechnology, and Molecular Biology. He was worked as Research assistant under an USDA project entitled "Molecular Detection of Bacterial and Fungal Diseases of Carp and Catfish and Herbal Treatment for Remedy of the Diseases". Now he is going to complete his Ph.D. degree at Auburn university. He has interest over microbiology, virology, cancer biology, and bioinformatics.

Jewel Halder
E-mail: halder98@gmail.com

Jewel Halder completed his post graduation in genetic engineering and biotechnology from Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh in 2009. He was worked as Research assistant under an USDA project entitled "Molecular Detection of Bacterial and Fungal Diseases of Carp and Catfish and Herbal Treatment for Remedy of the Diseases". He has interest over bioinformatics data analysis and structure prediction tools for drug designing, virology and cancer biology.

Raihan Mahmud
E-mail: bioraihan09@yahoo.com

Raihan Mahmud graduated from Shahjalal University of Science & Technology, Sylhet under the department of Genetic Engineering & Biotechnology in 2008. He was worked as Research assistant under an USDA project entitled "Molecular Detection of Bacterial and Fungal Diseases of Carp and Catfish and Herbal Treatment for Remedy of the Diseases". He has interest over bioinformatics data analysis and structure prediction tools for drug designing, virology and cancer biology.

Khairul Alam Mohammad Nayeem
E-mail: nayeem_bio@yahoo.com

Khairul Alam Mohammad Nayeem completed his post graduation in genetic engineering and biotechnology from Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh in 2009. He was worked as Research assistant under an USDA project entitled "Molecular Detection of Bacterial and Fungal Diseases of Carp and Catfish and Herbal Treatment for Remedy of the Diseases". He has interest over bioinformatics data analysis and structure prediction tools for drug designing, virology and cancer biology.

Nabangshu Shekhar Das
E-mail: nsdas10@gmail.com

Nabangshu Shekhar Das graduated from Shahjalal University of Science & Technology, Sylhet under the department of Genetic Engineering & Biotechnology in 2008. He was worked as Research assistant under an USDA project entitled "Molecular Detection of Bacterial and Fungal Diseases of Carp and Catfish and Herbal Treatment for Remedy of the Diseases". He has interest over virology, bioinformatics and drug designing, and cancer biology.

Muzahidul Islam Sarkar
E-mail: rokan_biotech@yahoo.com

Muzahidul Islam Sarkar completed his post graduation in genetic engineering and biotechnology from Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh in 2009. He was worked as Research assistant under an USDA project entitled "Molecular Detection of Bacterial and Fungal Diseases of Carp and Catfish and Herbal Treatment for Remedy of the Diseases". He has interest over bioinformatics, microbiology and cancer biology.

Janmajoy Dey
E-mail: deyjanmajoy@yahoo.com

Janmajoy Dey graduated from Shahjalal University of Science & Technology, Sylhet under the department of Genetic Engineering & Biotechnology in 2008. He was worked as Research assistant under an USDA project entitled "Molecular Detection of Bacterial and Fungal Diseases of Carp and Catfish and Herbal Treatment for Remedy of the Diseases". He has interest over bioinformatics, microbiology and drug designing.

Akhikun Nahar
E-mail: akhi.happy@yahoo.com

Akhikun Nahar graduated from Shahjalal University of Science & Technology, Sylhet under the department of Genetic Engineering & Biotechnology in 2008. She is pursuing research on "Purification and Molecular characterization of protease from *Serratia marcescens, Bacillus* sp. and *Pseudomonas* sp." She has interest over microbiology, virology and cancer biology, bioinformatics.