Physicochemical properties of insect and plant antifreeze proteins: a computational study

L. Ramya*

Centre for Nanotechnology and Advanced Biomaterials, SASTRA University, Thirumalaisamudram, Thanjavur 613 401, India

Antifreeze proteins are found in cold-surviving organisms. These proteins have greater structural diversity among same and different species. In this study, a total of 14 antifreeze proteins from both insects and plants were selected randomly and their physicochemical characteristics along with their structural features were analysed using computational tools. The results indicate that plant antifreeze proteins are mostly hydrophilic, which can interact with ice/water effectively. The study shows that the thermal stability of plant antifreeze proteins is greater than insect antifreeze proteins. Among the chosen sequences, insect antifreeze proteins were mostly β-sheet and plant antifreeze proteins were α-helix.

Keywords: Antifreeze proteins, disulphide bonds, homology modelling, hydrophobicity, thermal stability.

ANTIFREEZE proteins (AFPs) occur in polar fishes, insects, plants, fungi, bacteria and other organisms, and can inhibit ice growth at sub-zero temperatures. These proteins exhibit the properties of both thermal hysteresis (TH) and recrystallization inhibition (RI)[1]. Among the AFPs isolated from different species, fish AFPs are studied extensively from both experimental and computational approaches[2]. Fish AFPs are broadly classified as AFPs and antifreeze glycoproteins (AFGP)[1]. The physicochemical properties of fish AFPs were analysed previously[3,4] using bioinformatics tools. The complete biostatistics about the protein sequence which includes sequence length, physicochemical properties like molecular weight, extinction co-efficient, grand average hydropathy (GRAVY), aliphatic index (AI), instability index (II), etc. can be computed by computational tools and servers. From the literature it is known that fish AFPs show greater diversity in structure and each AFP represents distinct physicochemical and structural properties.

Insect AFPs have been characterized from both beetles (*Tenebrio molitor*[5], *Rhagium inquisitor*[6] and *Dendroides canadensis*[7]) and moth (*Choristoneura fumiferana*)[8]. Plant AFPs are found in winter rye[9], ryegrass[10] and carrot[11] which are also diverse in sequence. The structure and properties of these AFPs and the mechanism of inhibition of ice crystal growth have been reviewed in the literature[1,12–15]. Insect AFPs have greater TH (hyperactivity) which allows them to survive even at temperatures as low as 243 K (–30°C)[16]. Plant AFPs have several hydrophilic ice-binding domains which appear to function as inhibitors of ice recrystallization and ice nucleation[1]. AFPs in winter rye exhibit antifungal, hydrolytic activities apart from ice-binding activity[17]. These cold acclimated plants are resistant to injury caused by freezing[17]. AFPs from perennial ryegrass, carrot, and bittersweet nightshade are all glycosylated but bittersweet nightshade requires glycosylation for antifreeze activity[18].

Most of the plant AFPs are modified pathogenesis-related (PR) proteins that have high sequence identity and antifungal activities of the progenitor PR proteins[19]. Bansal *et al.*[20] studied the physicochemical properties, hydropathicity and secondary structure of seven different AFPs of *Typhula ishikariensis* (a plant pathogen). Since numerous experimental and computational studies were performed on fish AFPs and little is known about insect and plant AFPs, the main focus here is to study the properties of insect and plant AFPs. In order to know the similarities and differences among insect and plant AFPs, their physicochemical characterization was studied using computational tools.

Methods and materials

Selection of AFP sequence

Insect and plant AFP sequences were retrieved from the protein database Swiss-Prot[21]. An advanced search tool was used to filter for insect and plant AFPs with sequence length less than 300. From the list, a total of 14 AFP sequences (6 insect AFPs and 8 plant AFPs) were considered for the study from different organisms with no redundancy of the data set. Table 1 gives the AFP sequences chosen for the study. FASTA format of the sequence was used for the analyses.

Software used to study the physicohemical properties of AFPs

ExPASy's ProtParam (http://us.expasy.org/tools/protparam.html)[22] was used to study the characteristics of

*e-mail: lramya174@gmail.com

Table 1. Insect (*) and plant antifreeze protein (**) sequences retrieved from Swiss-Prot database for study

Accession number	Sequence description	Organism
O16119*	Insect antifreeze protein	*Tenebrio molitor* (yellow mealworm beetle)
O46351*	Insect antifreeze protein	*Dendroides canadensis*
E5LR38*	Insect antifreeze protein	*Rhagium inquisitor*
A1IIC7*	Insect antifreeze protein	*Dorcus curvidens binodulosus*
F2VKG6*	Insect antifreeze protein	*Anatolica polita*
Q9GSA6*	Insect antifreeze protein	*Choristoneura fumiferana*
Q9AXR9**	31.7 kDa class I endochitinase-antifreeze protein	*Secale cereale*
Q9AXR8**	24.8 kDa class II endochitinase-antifreeze protein	*Secale cereale*
Q8S5Z3**	Antifreeze glycoprotein	*Oryza sativa* subsp. *Japonica* (rice)
Q84LQ7**	29 kDa chitinase-like thermal hysteresis protein	*Solanum dulcamara* (climbing nightshade)
Q42390**	Daucus carota globulin-like protein	*Daucus carota* (wild carrot)
B5T007**	Ice recrystallization inhibition protein-like protein	*Lolium perenne* (*Perennial ryegrass*)
Q9S9D9**	AFA3 = Antifreeze protein	*Nicotiana tabacum* (common tobacco)
Q9S899**	18.4 kDa candidate antifreeze protein	*Pinus monticola* (western white pine) (*Strobus monticola*)

physicochemical properties of both insect and plant AFPs. Table 2 gives the percentage of amino acids composition for each of the protein sequences. The physicochemical properties like molecular weight, theoretical isoelectric point (p^1), total number of positive and negative residues, extinction coefficient[23], instability index[24], aliphatic index[25] and grand average hydrophathy (GRAVY)[26] were computed for these proteins. The secondary structural details were studied using SOPMA (self-optimized prediction method and alignment)[27]. The transmembrane regions in proteins were identified using SOSUI server[28]. The presence of numbers of cysteine and those involved in disulphide bonds was found using DiANNA (diamino acid neural network application) 1.1 web server[29]. When the experimental three-dimensional structures of proteins are unavailable it becomes essential to perform homology modelling to study the conformational features. The modelling of three-dimensional structure of these AFPs was predicted using the Swiss-Model program[30].

Results and discussion

From the percentage of occurrence of amino acids (Table 2), it was clear that most of the selected AFPs lack the significant presence of aromatic residues (Phe, Trp and Tyr). It was evident from the existence of Ser and Thr, that these AFPs were mostly hydrophilic. The presence of Cys was higher in insect AFPs (except for E5LR38, *R. inquisitor*) compared to that of plant AFPs which effectively form disulphide bonds.

Physicochemical properties

Table 3 gives the results of primary structural analysis of protein sequences. The isoelectric point (p^1) is the value at which the molecule carries no charges or the positive and negative charges are equal. If the computed p^1 values are less than 7, then the AFPs are acidic, otherwise they are basic in nature. Among insect AFPs, three of them were basic and other three sequences were acidic. Thus insect AFPs cannot be categorized as acidic or basic from their respective p^1 values. Plant AFPs are found to be preferably basic among the selected sequences. The two plant AFPs namely Q9S9D9 (*Nicotiana tabacum*) and Q9S899 (*Pinus monticola*) were acidic which are actually very small polypeptides with residues 38 and 22. Studies showed that fish AFPs were basic with an average p^1 value of 6.73 (refs 3, 4). Thus to purify a protein by isoelectric focussing method, the p^1 value of the protein will be useful in developing buffer systems[3].

Extinction coefficient (EC) values help in the quantitative study of protein and protein–ligand interactions in solution. EC are in units of M^{-1} cm^{-1}, at 280 nm measured in water. EC of AFPs at 280 nm ranges from 125 to 14,325 M^{-1} cm^{-1} for insect AFPs and from 5500 to 56,350 M^{-1} cm^{-1} for plant AFPs with respect to the concentration of Cys, Trp and Tyr. The high EC value of plant AFPs Q9AXR9 (*Secale cereale*) followed by Q42390 (*Daucus carota*) indicates the presence of high concentration of Cys, Trp and Tyr. ExPasy Protparam computes no EC value for plant AFP Q9S9D9 (*N. tabacum*) due to the absence of Cys, Trp and Tyr. This indicates that this plant AFP (Q9S9D9, *N. tabacum*) cannot be analysed using UV spectral methods.

The bio-computed half-life of AFPs was greater than 20 h except for plant AFP Q9AXR9 (*S. cereale*) which has only 30 min. A protein with instability index (II) smaller than 40 was predicted as stable and the one above 40 was unstable[24]. The II value ranges from 11.47 to 79.09 for insect AFPs and from 1.31 to 77.00 for plant AFPs. Most of the insect AFPs were predicted to be unstable except for E5LR38 (*R. inquisitor*) and Q9GSA6 (*C. fumiferana)*. Plant AFPs were found to be highly stable except Q9S899 (*P. monticola*). The plant AFP *N. tabacum* (Q9S9D9) was found to be highly stable with lowest instability index of 1.31.

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

Amino acids	O16119*	O46351*	E5LR38*	A1IIC7*	F2VKG6*	Q9GSA6*	Q9AXR9**	Q9AXR8**	Q8S5Z3**	Q84LQ7**	Q42390**	B5T007**	Q9S9D9**	Q9S899**
Ala	9.5	6.5	19.4	9.5	9.4	6.6	9.4	11.2	34.5	6.7	3.7	6.7	60.5	9.1
Arg	0.0	3.7	5.2	0.7	3.8	2.5	4.7	6.4	6.8	5.2	3.3	4.3	2.6	9.1
Asn	10.7	7.4	1.5	8.0	9.4	6.6	4.7	6.0	0.7	7.1	4.8	12.2	5.3	4.5
Asp	4.8	4.6	1.5	3.6	3.8	5.0	5.0	6.0	1.4	4.5	5.0	4.7	7.9	0.0
Cys	19.0	17.6	1.5	16.1	17.0	8.3	5.7	2.1	1.4	4.1	0.2	2.4	0.0	0.0
Gln	3.6	1.9	4.5	2.9	2.8	5.0	4.7	3.4	2.7	4.5	3.1	1.6	0.0	4.5
Glu	1.2	2.8	0.7	1.5	0.9	0.0	2.0	2.1	0.7	1.5	7.3	1.6	0.0	13.6
Gly	9.5	4.6	14.9	4.4	2.8	6.6	13.1	12.0	1.4	13.9	8.5	12.2	0.0	4.5
His	2.4	1.9	0.7	0.7	0.9	0.0	1.3	0.9	0.7	1.9	2.5	3.5	0.0	0.0
Ile	0.0	2.8	2.2	2.2	1.9	8.3	3.7	3.9	1.4	5.2	6.2	2.4	0.0	0.0
Leu	0.0	2.8	0.0	3.6	2.8	2.5	4.7	4.3	4.7	4.9	8.3	9.4	5.3	4.5
Lys	3.6	2.8	1.5	2.2	1.9	4.1	2.7	2.1	0.7	3.4	7.7	2.0	2.6	0.0
Met	0.0	1.9	0.7	1.5	2.8	0.8	0.7	2.6	10.1	1.5	2.5	1.6	2.6	0.0
Phe	1.2	0.0	1.5	2.2	1.9	0.0	4.7	4.7	0.7	6.0	6.7	1.6	0.0	0.0
Pro	2.4	1.9	2.2	1.5	1.9	1.7	7.0	4.3	7.4	6.4	4.2	2.8	0.0	0.0
Ser	8.3	10.2	11.9	9.5	6.6	15.7	8.4	6.4	3.4	7.9	10.0	12.2	2.6	13.6
Thr	20.2	14.8	26.9	21.2	21.7	16.5	6.4	9.4	7.4	6.0	4.4	7.1	10.5	13.6
Trp	0.0	0.9	0.0	0.7	0.9	0.0	2.0	1.7	4.1	2.2	1.2	1.6	0.0	4.5
Tyr	1.2	4.6	0.0	3.6	4.7	4.1	5.0	5.2	0.0	3.7	2.9	0.4	0.0	0.0
Val	2.4	6.5	3.0	4.4	1.9	5.8	4.0	5.2	10.1	3.4	7.5	9.8	0.0	18.2

*Insect AFPs. **Plant AFPs.

Table 3. Physicochemical properties of both insects and plants AFPs computed using ExPASy's ProtParam tool.

Accession number	Length	Molecular weight	Isoelectric point	Negative R group	Positive R group	Extinction coefficient (M^{-1} cm^{-1})	Instability index	Aliphatic index	GRAVY
O16119*	84	8384.0	5.30	5	3	2490	49.98	16.43	–0.443
O46351*	108	11583.0	6.00	8	7	14075	79.09	46.94	–0.125
E5LR38*	134	12542.4	11.28	3	9	125	11.47	36.79	–0.315
A1IIC7*	137	14296.9	4.59	7	4	14325	47.27	44.96	0.060
F2VKG6*	106	11472.8	7.31	5	6	14075	51.14	33.30	–0.211
Q9GSA6*	121	12538.9	8.09	6	8	8075	26.96	65.29	–0.149
Q9AXR9**	298	31683.2	7.57	21	22	56350	44.32	53.79	–0.325
Q9AXR8**	233	25066.8	7.81	19	20	40130	33.58	57.90	–0.345
Q8S5Z3**	148	15130.1	11.83	3	11	33125	36.93	87.57	0.751
Q84LQ7**	267	28845.2	8.88	16	23	48525	38.14	55.96	–0.391
Q42390**	481	53875.2	6.10	59	53	53860	34.99	82.20	–0.278
B5T007**	254	26248.0	7.11	16	16	23865	26.31	81.30	–0.215
Q9S9D9**	38	3359.6	4.43	3	2	Nil	1.31	81.05	0.563
Q9S899**	22	2419.6	4.79	3	2	5500	77.00	79.55	–0.368

*Insect AFPs; **Plant AFPs. GRAVY, Grand average hydropathy.

AI is defined as the relative volume of a protein with aliphatic side chains (A, V, I and L). This was regarded as the positive factor for increase of thermal stability of globular proteins. The AI value varies from 16.43 to 65.29 for insect AFPs and from 53.79 to 87.57 for plant AFPs. Plant AFPs with AI values higher than insect AFPs indicate that they are most stable at a wide range of temperatures. The lower value of AI indicates that the proteins are more flexible. Very high values of AI for fish AFPs were all found to be type III proteins[4]. Thus the order of greater thermal stability among AFPs will be fish type III AFPs > fish type I/II > AFPs plant AFPs > insect AFPs.

The GRAVY value for a protein was calculated as a sum of hydropathy value of all amino acids, divided by the number of residues in the sequences. GRAVY index ranges from –0.443 to +0.060 for insect AFPs and from –0.391 to +0.751 for plant AFPs. GRAVY values for fish AFPs were observed to be positive[3,4]. From Table 3 it is clear that except for A1IIC7 (*Dorcus curvidens binodulosus*) all other insect AFPs have negative GRAVY values. Similarly among plant AFPs, Q8S5Z3 (*Oryza sativa*

Table 4. Transmembrane regions identified by SOSUI server

Accession number	Transmembrane region (N terminal – C terminal)	Type	Length
O46351*	MVWVCKNSILVISVVLMYVCHEC (1–23)	Primary	23
A1IIC7*	TCAFTKSWLVVAVIVMCLCTGY (5–26)	Primary	22
F2VKG6*	MALTTKWFLIAVIVMCLCAEYYC (1–23)	Primary	23
Q8S5Z3**	TAAAWICSAVVAAAR (23–37)	Secondary	15
	TAAPMQFVVAMATRIWSAAVAAS (44–66)	Primary	23
	AVAAAPMLPAATAAAAMLPAAAA (82–104)	Primary	23
	LPATATAAPMQAVTVSVVWMCLA (118–140)	Primary	23
Q42390**	MGKGLTFLLVLVLVISVK (1–18)	Primary	18
B5T007**	KCLMLLLSFAFLLSAAGTATATP (3–25)	Primary	23
	VTALWLPRSGLTGPIPSWICQLH (73–95)	Secondary	23

* Insect AFPs; **Plant AFPs.

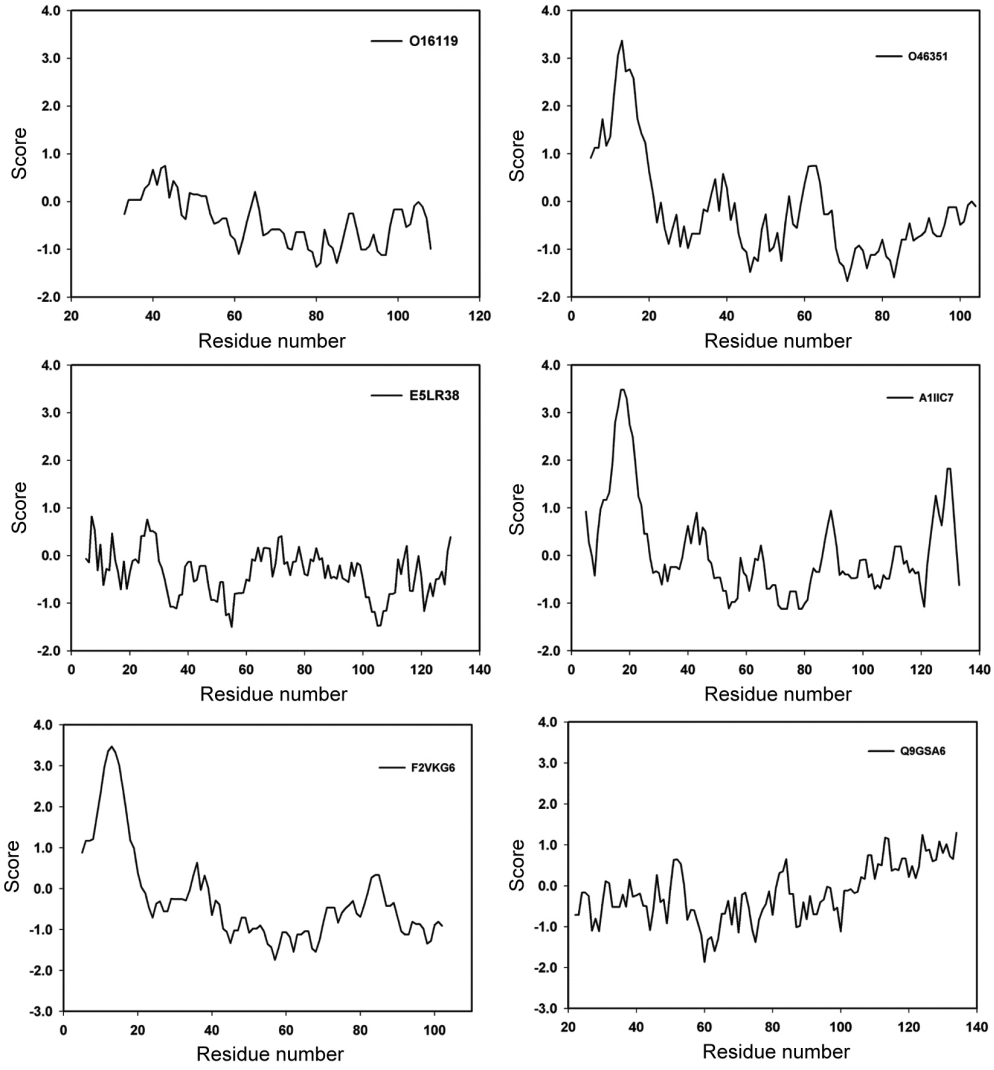

Figure 1. Kyte and Doolittle hydrophobicity profile computed for the insect antifreeze protein sequences.

subsp. *Japonica*) has high positive GRAVY value of +0.751 followed by Q9S9D9 (*N. tabacum*) with +0.563. Thus most of the chosen insect and plant proteins have low negative values of GRAVY which indicate their better interaction with water molecules.

Transmembrane regions were identified to understand the functional characterization of AFPs. The SOSUI server identified the transmembrane regions with their length and differentiated membrane proteins from soluble proteins. SOSUI classified O16119 (*T. molitor*), E5LR38

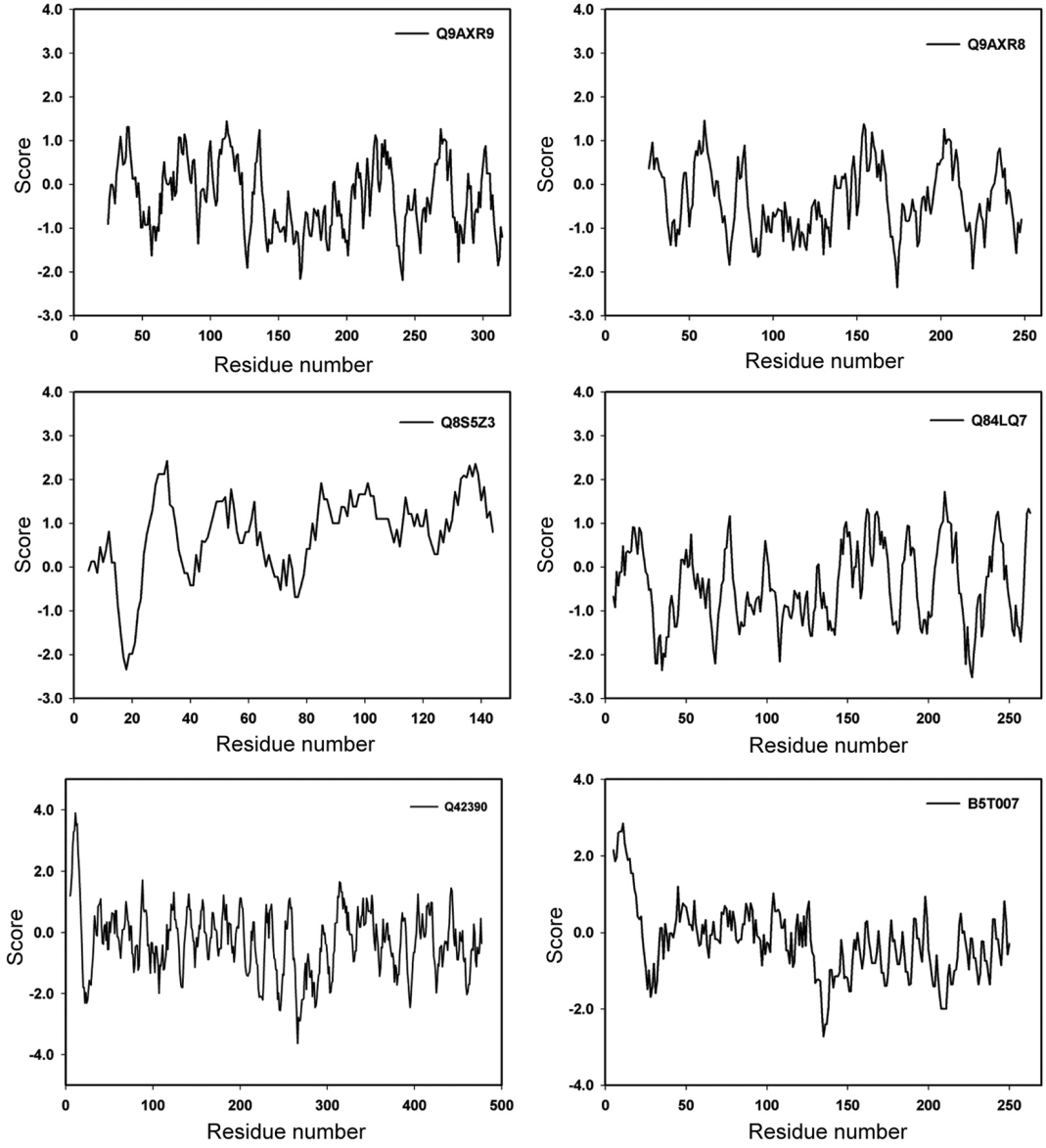

Figure 2. Kyte and Doolittle hydrophobicity profile computed for the plant antifreeze protein sequences.

(*R. inquisitor*) and Q9GSA6 (*C. fumiferana*) as soluble proteins and the rest of insect AFPs as membrane proteins. The server identified one transmembrane region in each of the three insect AFPs O46351, A1IIC7 and F2VKG6. The transmembrane regions and their sequence lengths are tabulated in Table 4. Similarly in the case of plant AFPs, three proteins namely Q8S5Z3 (*O. sativa* subsp. *Japonica*), Q42390 (*D. carota*) and B5T007 (*Lolium perenne*) were membrane proteins and the remaining were soluble proteins. There were four transmembrane regions in Q8S5Z3, one in Q42390 and two in B5T007. Mutation studies showed that hydrophobic interactions drive the interaction between protein and ice crystal and not the hydrogen bonds[31]. From literature survey[3,4], it was observed that most of the fish AFPs were hydrophobic in nature whereas plant AFPs were highly soluble. This indicates that plant AFPs can have strong interaction with water molecules and thereby block the interaction of waters with ice crystal. Here the plant AFP Q8S5Z3 (*O. sativa* subsp. *Japonica*) with GRAVY value of 0.751 is highly hydrophobic with four transmembrane regions. But the protein sequence with highest AI of 87.57 shows stability at extreme temperatures. Thus the physicochemical properties make it curious to study the interaction of Q8S5Z3 (*O. sativa* subsp. *Japonica*) with ice crystal.

Kyte and Doolittle hydrophobicity plot[26] for both insect and plant AFP sequences with window size 9 residues, indicates that the transmembrane regions are rich in hydrophobic amino acids (Table 4). For insect AFPs

Table 5. Percentage occurrence of secondary structural features in each protein sequences calculated using SOPMA web server

Accession number	Percentage occurrence of secondary structural features			
	α-helix	Extended	β-turn	Random coil
O16119*	0.00	23.81	1.19	75.00
O46351*	15.74	17.59	7.41	59.26
E5LR38*	11.94	27.61	11.94	48.51
A1IIC7*	13.87	35.04	5.84	45.26
F2VKG6*	11.32	28.30	1.89	58.49
Q9GSA6*	13.22	30.58	7.44	48.76
Q9AXR9**	23.49	16.44	7.05	53.02
Q9AXR8**	26.61	17.60	9.87	45.92
Q8S5Z3**	66.89	6.76	0.68	25.68
Q84LQ7**	20.22	20.97	7.87	50.94
Q42390**	19.96	31.60	13.93	34.51
B5T007**	18.90	24.02	12.60	44.49
Q9S9D9**	92.11	0.00	0.00	7.89
Q9S899**	72.73	4.55	0.00	22.72

*Insect AFPs; **Plant AFPs.

Table 6. Disulphide (SS) bonds predicted by DiANNA web server

Accession number	No. of Cys residues	Residues involved in disulphide bonds
O16119*	16	2–15, 8–18, 11–21, 27–45, 33–39, 51–75, 57–63, 69–81
O46351*	19	5–83, 20–23, 27–95, 33–40, 36–107, 46–59, 53–89, 65–71, 77–101
E5LR38*	2	Nil
A1IIC7*	22	6–21, 23–61, 28–46, 30–121, 36–43, 39–97, 49–55, 67–115, 73–79, 85–91, 103–109
F2VKG6*	18	16–63, 18–105, 23–45, 26–35, 32–87, 39–93, 42–51, 57–81, 69–75
Q9GSA6*	11	4–38, 17–116, 25–98, 56–93, 68–111
Q9AXR9**	17	3–24, 12–15, 17–18, 31–42, 35–39, 78–140, 152–160, 259–291
Q9AXR8**	5	26–212, 193–225
Q8S5Z3**	2	29–138
Q84LQ7**	11	3–184, 7–101, 113–121, 144–239, 220–252
Q42390**	1	Nil
B5T007**	6	4–92, 26–66, 58–59
Q9S9D9**	0	Nil
Q9S899**	0	Nil

*Insect AFPs; **Plant AFPs.

(O46351, A1IIC7 and F2VKG6), the plot shows a high positive score for transmembrane regions (Figure 1). Similarly for plant AFPs Q8S5Z3 and Q42390, the plot shows a high positive score for transmembrane regions (Figure 2).

The secondary structural features of protein sequences were predicted using SOPMA (self-optimized prediction method and alignment)[27]. The analyses point out whether the amino acid falls in the region of helix, strand or coil. The percentage occurrences of these details are tabulated in Table 5. Occurrence of random coils dominated among secondary structures followed by β-strands in insect AFPs. Three plant AFPs namely Q8S5Z3 (*O. sativa* subsp. *Japonica*), Q9S9D9 (*N. tabacum*) and Q9S899 (*P. monticola*) were found to be rich in α-helix whereas in other AFPs random coils were dominant followed by α-helix and β-strands. Other important secondary structures like 3_{10} helix, π-helix, bend and β-bridgeswere not present. For fish AFPs[3], 10 out of 15 sequences were found to be α-helix followed by random coils, thus indicating the α-helix was the dominating secondary structural feature. Proline has a property of creating kinks and bends in polypeptide chain and is usually a helix breaker. Interestingly, plant AFP Q8S5Z3 (*O. sativa* subsp. *Japonica*) which has the highest percentage occurrence of proline (7.4%) is found to be rich in α-helix (66.89%).

The DiANNA web server[29] identifies the presence of S–S bonds among all Cys residues in the 14 protein sequences and are tabulated in Table 6. The Cys residues present in insect AFPs were all involved in S–S bond except for E5LR38 (*R. inquisitor*) which had no S–S

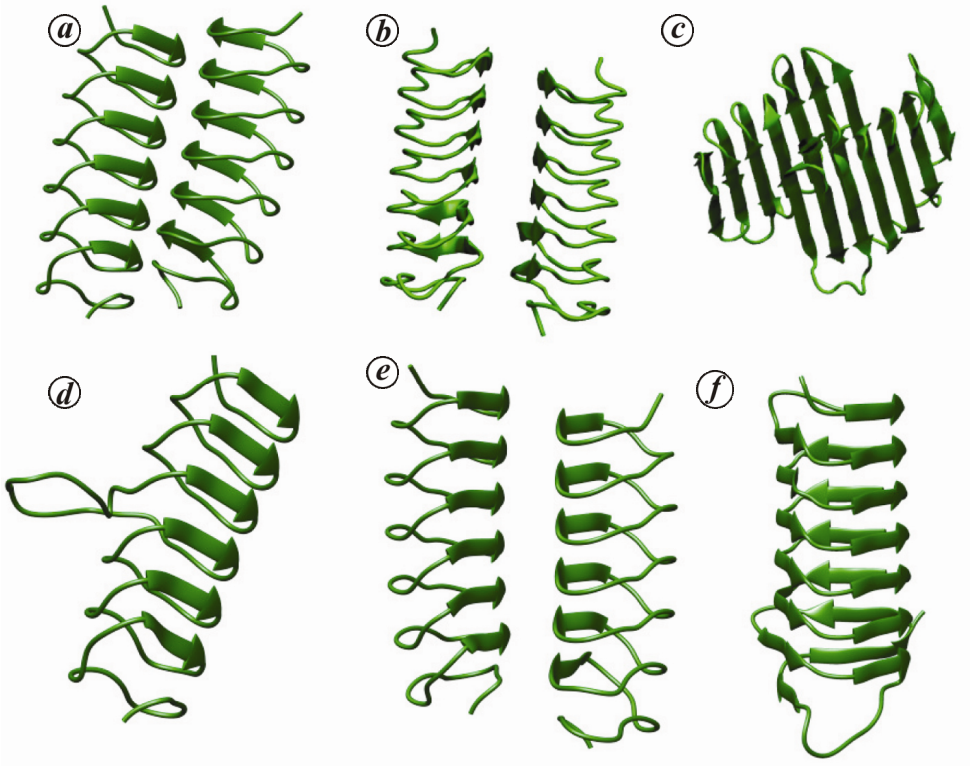

Figure 3. The best modelled structure for insect AFPs from Swiss-Model with more than 90% of residues in the favoured regions as predicted by RAMPAGE server. ***a***, *Tenebrio molitor* (O16119); ***b***, *Dendroides canadensis* (O46351); ***c***, *Rhagium inquisitor* (E5LR38); ***d***, *Dorcus curvidens binodulosus* (A1IIC7); ***e***, *Anatolica polita* (F2VKG6); ***f***, *Choristoneura fumiferana* (Q9GSA6).

bond with the presence of just two Cys residues. There were few S–S bonds present among plant AFPs as the occurrence of Cys is limited in those sequences.

Homology modelling and validation of modelled structures

Three-dimensional protein structure explains the molecular basis of protein function. It is difficult to crystallize membrane proteins and here we have three sequences in each of insect and plant AFPs as membrane proteins. Since we have studied the physicochemical properties of these proteins, it will be useful to design experiments to study their 3D structures. The structures modelled using Swiss-Model were observed to be most acceptable[4]. Two plant AFPs Q8S5Z3 (*O. sativa* subsp. *Japonica*) and Q9S899 (*P. monticola*) were not modelled by the program stating template non-availability for the first protein and residues less than 30 for the second protein. The Swiss-Model template library was searched with BLAST[32] and HHBlits[33] for evolutionary related structures matching the target sequence. The Swiss-Model program modelled three structures for each of the protein sequence from several varying number of template structures.

The modelled 3D structures were evaluated using the online server RAMPAGE[34]. An evaluation of residues in different favoured, allowed and outlier regions can reveal the quality of the modelled structure. The highest percentage of residues in the allowed region predicted was considered the best modelled structure. Figures 3 and 4 show the best modelled structure for both insect and plant AFPs. Insect AFPs were found to be β-sheets. The AFPs O16119 (Figure 1 *a*), 046351 (Figure 1 *b*) and F2VKG6 (Figure 1 *e*) seem to have common β-helix structure with residue regions 14–16, 26–28, 38–40, 50–52, 62–64 and 74–76 for O16119 (*Tenebrio molitor*), 39–41, 51–54, 64–66, 76–78, 88–90 and 100–102 for 046351 (*D. canadensis*) and 38–40, 51–52, 62–64, 74–76, 86–88 and 98–100 for F2VKG6 (*Anatolica polita*). Here F2VKG6 (*A. polita*) was a membrane protein with residues in the transmembrane region 1–23 (Table 4). All these proteins were modelled using the same protein template 1EZG.pdb. Plant AFPs were α-helix among the chosen protein sequences as predicted by SOPMA server. Table 5 shows that AFPs almost have equal presence of α-helix and β-sheets except for Q9S9D9. Figure 4 shows the presence of both secondary structural regions (α-helix and β-sheets) as predicted by the Swiss-Model program. The predicted structures of both insect and plant AFPs were further validated using PROCHECK program[35] and the φ–ψ plot analyses were tabulated in Table S1 (see Supplementary Information, online).

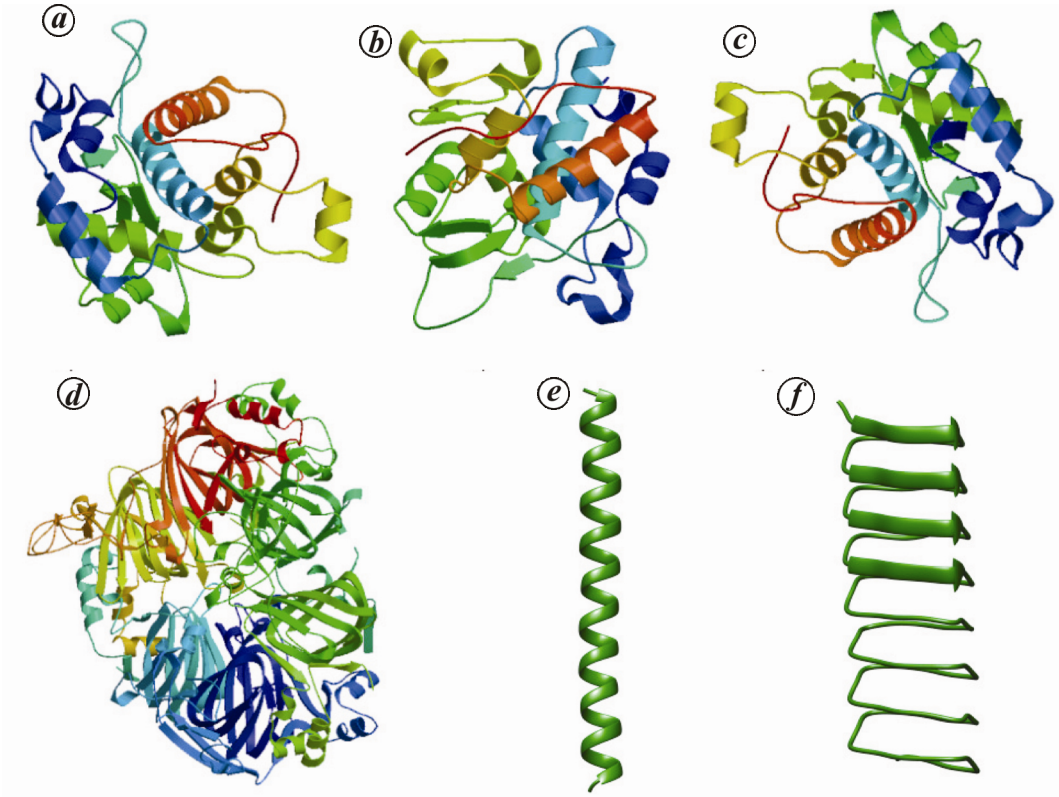

Figure 4. The best modelled structure for plant AFPs from Swiss-Model with more than 90% of residues in the favoured regions as predicted by RAMPAGE server. ***a***, *Secale cereale* (Q9AXR9); ***b***, *Secale cereale* (Q9AXR8); ***c***, *Solanum dulcamara* (Q84LQ7); ***d***, *Daucus carota* (Q42390); ***e***, *Nicotiana tabacum* (Q9S9D9); ***f***, *Lolium perenne* (B5T007).

Conclusion

The study of physicochemical properties gives an overall view about the insect and plant AFPs. The insect AFPs have more Cys residues and thereby form a larger number of S–S bonds than plant AFPs. Plant AFPs were preferably hydrophilic unlike fish AFPs and so these proteins can interact with ice/water molecules more effectively and inhibit the further growth of ice crystal. Since it was known that plant AFPs were weak in their TH activity[36], the study of interaction between plant AFPs and ice crystal becomes essential to understand their biological activity.

From the study of aliphatic index, the order of greater thermal stability among these proteins will be fish type III AFPs > fish type I/II AFPs > plant AFPs > insect AFPs. Thus next to fish AFPs, plant proteins can withstand wide range of temperatures. Unlike fish AFPs, most of the proteins (both from insect and plant of this study) have low negative values of GRAVY indicating their better interaction with water molecules. One unusual observation was the plant AFP Q8S5Z3 (*O. sativa* subsp. *Japonica*) found to be rich in α-helix (66.89%) with highest percentage of proline presence (7.4%). It was a membrane protein with GRAVY 0.751 (highest value of the chosen AFPs) and has higher p^{I} value (11.83) representing basic nature. Thus the study shows the need to understand the interaction of AFPs with water/ice molecules particularly plant AFP Q8S5Z3 (*O. sativa* subsp. *Japonica*).

Competing interests: The author declares that there is no competing interest.

1. Jia, Z. and Davies, P. L., Antifreeze proteins: an unusual receptor–ligand interaction. *Trends Biochem. Sci*., 2002, **27**, 101–106.
2. Venketesh, S. and Dayananda, C., Properties, potentials, and prospects of antifreeze proteins. *Crit. Rev. Biotechnol*., 2008, **28**, 57–82.
3. Sivakumar, K., Balaji, S. and Gangaradhakrishnan, In silico characterization of antifreeze proteins using computational tools and servers. *J. Chem. Sci*., 2007, **119**, 571–579.
4. Md. Hossain, M., Fish antifreeze proteins: computational analysis and physicochemical characterization. *Int. Curr. Pharm. J*., 2012, **1**, 18–26.
5. Liou, Y. C., Tocilj, A., Davies, P. L. and Jia, Z., Mimicry of ice structure by surface hydroxyls and water of a beta-helix antifreeze protein. *Nature*, 2000, **406**, 322–324.
6. Kristiansen, E., Ramlov, H., Hagen, L., Pedersen, S. A., Andersen, R. A. and Zachariassen, K. E., Isolation and characterization of hemolymph antifreeze proteins from larvae of the longhorn beetle Rhagium inquisitor (L.). *Comp. Biochem. Physiol. B Biochem. Mol. Biol*., 2005, **142**, 90–97.
7. Andorfer, C. A. and Duman, J. G., Isolation and characterization of cDNA clones encoding antifreeze protein of the pyrochorid

beetle *Dendrodies canadensis*. *J. Insect Physiol*., 2000, **46**, 365–372.
8. Tyshenko, M. G., Doucet, D., Davies, P. L. and Walker, V. K., The antifreeze potential of the spruce budworm thermal hysteresis protein. *Nat. Biotechnol.*, 1997, **15**, 887–890.
9. Hon, W. C., Griffith, M., Mlynarz, A., Kwok, Y. C. and Yang, D. S. C., Antifreeze proteins in winter rye are similar to pathogenesis-related proteins. *Plant Physiol*., 1995, **109**, 879–889.
10. Sidebottom, C. *et al.*, Heat-stable antifreeze protein from grass. *Nature*, 2000, **406**, 256–256.
11. Worrall, D. *et al.*, A carrot leucine-rich repeat protein that inhibits ice recrystallization. *Science*, 1998, **282**, 115–117.
12. Davies, P. L. and Sykes, B. D., Antifreeze proteins. *Curr. Opin. Struct. Biol*., 1997, **7**, 828–834.
13. Ewart, K. V., Lin, Q. and Hew, C. L., Structure, function and evolution of antifreeze proteins. *Cell Mol. Life. Sci*., 1999, **55**, 271–283.
14. Ben, R. N., Antifreeze glycoproteins – preventing the growth of ice. *Chem. Bio. Chem*., 2001, **2**, 161–166.
15. Tachibana, Y., Fletcher, G. L., Fujitani, N., Tsuda, S., Monde, K. and Nishimura, S. I., Antifreeze glycoproteins: Elucidation of the structural motifs that is essential for antifreeze activity. *Angew. Chem. Int*., 2004, **43**, 856–862.
16. Duman, J. G., Antifreeze and ice nucleator proteins in terrestrial arthropods. *Annu. Rev. Physiol*., 2001, **63**, 327–357.
17. Hiilovaara-Teijo, M., Hannukkala, A., Griffith, M., Yu, X.-M., and Pihakaski-Maunsbach, K., Snow-mold-induced apoplastic proteins in winter rye leaves lack antifreeze activity. *Plant Physiol*., 1999, **121**, 665–673.
18. Duman, J. G., Purification and characterization of a thermal hysteresis protein from a plant, the bittersweet nightshade *Solanum dulcamara*. *Biochim. Biophys. Acta*, 1994, **1206**, 129–135.
19. Griffith, M. and Yaish, W. F., Antifreeze proteins in overwintering plants: a tale of two activities. *Trends Plant. Sci*., 2004, **9**, 399–405.
20. Bansal, H., Narang, D. and Jabalia, N., Computational characterization of antifreeze proteins of *typhula ishikariensis* – gray snow mould. *J. Proteins Proteomics*, 2014, **5**, 169–176.
21. Boeckmann, B. *et al.*, The SWISS-PROT protein knowledgebase and its supplement TrEMBL. *Nucleic Acids Res*., 2003, **31**, 365–370.
22. Gasteiger, E., Protein identification and analysis tools on the ExPASy Server. In *The Proteomics Protocols Handbook* (ed. Walker, J. M.), Humana Press, 2005, pp. 571–607.
23. Gill, S. C. and Von Hippel, P. H., Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem*., 1989, **182**, 319–328.
24. Guruprasad, K., Reddy, B. V. P. and Pandit, M. W., 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 Eng. Des. Sel.*, 1990, **4**, 155–164.
25. Ikai, A. J., Thermo stability and aliphatic index of globular proteins. *J. Biochem*., 1980, **88**, 1895–1898.
26. Kyte, J. and Doolittle, R. F., A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.,* 1982, **157**, 105–132.
27. Geourjon, C. and Deléage, G., SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.*, 1995, **11**, 681–684.
28. Hirokawa, T., Boon-Chieng, S. and Mitaku, S., SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, 1998, **14**, 378–379.
29. Ferre, F. and Clote, P., DiANNA: a web server for disulfide connectivity prediction. *Nucleic Acids Res*., 2005, **33** (web server issue), W230–W232.
30. Arnold, K., Bordoli, L., Kopp, J. and Schwede, T., The Swiss–Model workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 2006, **22**, 195–201.
31. Haymet, A. D. J., Ward, L. G. and Harding, M. M., Winter Flounder 'Antifreeze' proteins: Synthesis and ice growth inhibition of analogues that probe the relative importance of hydrophobic and hydrogen bonding interactions. *J. Am. Chem. Soc*., 1999, **121**, 941–948.
32. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*., 1997, **25**, 3389–3402.
33. Remmert, M., Biegert, A., Hauser, A. and Soding, J., HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods*, 2012, **9**, 173–175.
34. Lovell, S. C. *et al.*, Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Struct, Funct, Bioinf*., 2003, **50**, 437–450.
35. Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M., PROCHECK – a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, 1993, **26**, 283–291.
36. Urrutia, M. E., Duman, J. G. and Knight, C. A., Plant thermal hysteresis proteins. *Biochim. Biophys. Acta*, 1992, **1121**, 199–206.

ACKNOWLEDGMENTS. The present work was supported and funded by the Prof. T. R. Rajagopalan fund from SASTRA University.

Received 7 March 2016; accepted 18 October 2016

doi: 10.18520/cs/v112/i07/1512-1520