Journal of Bioinformatics and Sequence Analysis Vol. 3(4), pp. 70-88, May 2011
Available online at http://www.academicjournals.org/JBSA
DOI: 10.5897/JBSA11.003
ISSN 2141-2464 ©2011 Academic Journals

*Full length Research Paper*

# A framework for classification of antifreeze proteins in over wintering plants based on their sequence and structural features

**J. Muthukumaran[1], P. Manivel[1], M. Kannan[1], J. Jeyakanthan[2] and R. Krishna[1]***

[1]Centre for Bioinformatics, School of Life Sciences, Pondicherry University, Puducherry – 605 014, India.
[2]Department of Bioinformatics, Alagappa University, Karaikudi, 630 003, India.

Accepted 11 April, 2011

**Overwintering plants produce antifreeze proteins (AFPs) which permits the plant survival in cold condition. Analysis of sequence and structural features of these proteins would help in better understanding of their functions. In this study, we report the analysis of 40 plant AFPs on the basis of sequence and structural based classification scheme (CS). Sequence based CS segregates the AFPs into various categories such as physicochemical properties, transmembrane regions, glycosylation sites, and sub cellular localization. Phylogeny based CS separate the chosen proteins into several groups, in which, the AFP from *Festuca pratensis*, *Pinus monticola*, *Ricinus communis* and *Populus suaveolens* are newly identified leucine rich repeat (LRR), pathogenesis related (PR), hemagglutinin related (HR) and pleckstrin homology (PH) family, respectively. The secondary and 3D structures of 27 AFPs were predicted, whereas the remaining 13 protein structures were reported in different studies. Selected proteins are found to have mixed secondary structural elements and the more coil like content were observed in few of the proteins. The proposed classification scheme in over wintering plants can be useful in searching the newly sequenced plant genome for putative AFPs or designing an engineered construct helpful for several industrial and biomedical applications.**

**Key words:** Antifreeze protein, ice-structuring protein, ice-binding protein, thermal hysteresis proteins, over wintering plants.

## INTRODUCTION

Antifreeze proteins (AFPs) have an affinity for ice due to their structural complementarity nature, thereby inhibiting the ice-crystal growth. It protects organisms from deep freezing temperatures and is expressed in vertebrates, invertebrates, plants, bacteria, and fungi (Venketesh and Dayananda, 2008). Adsorption of AFPs onto ice surfaces has two different effects such as thermal hysteresis (TH) and recrystallization inhibition (RI). These two properties of AFPs, which prevents the growth of ice crystals, can be used in the development of transgenic plants with antifreeze properties thereby the yield of important crops can be increased. The activity of AFPs can be quantitatively assayed by measuring the TH activity, and it is also qualitatively assayed by examining the morphology of ice crystals grown in the AFPs. They do not prevent the ice formation, but instead function by changing the morphology of ice crystal and inhibit its further escalation (Griffith and Yaish, 2004). AFPs are used to preserve the cells, tissues and organs for transplant or transfusion in medicine at low temperature. It is also used to improve the production of farm fishes in winter season. There is also rising evidence that AFPs interact with mammalian cell membranes to protect from freezing damage through cold acclimatization (Fletcher et al., 2001).

The interaction of AFPs with ice crystals is a precise process, and it is mediated through non covalent interactions between hydrophilic groups of the AFP and oxygen atoms of ice lattice (Bayer-Giraldi et al., 2010).

AFPs can be classified into five different types in which

*Corresponding author. E-mail: krishstrucbio@gmail.com, krishna.bic@pondiuni.edu.in. Tel: +91-413-2655580. Fax: +91-413-2655211.

Type I, Type I-hyp, II, III and IV are belonging to fishes and type V AFPs are hyperactive (greater TH value), found in insects. Type I AFPs are found in flounder and sculpin, which is amphiphilic and single α- helical structure containing ~30 to 50 residues with putative ice-binding threonine residues, and it is repeated every 11 residues along the length of the helix (Sicheri and Yang, 1995). Type I-hyp AFPs are found in many right eye flounders, which is ~32 kD and it is substantially superior at depressing the freezing temperature than most fish AFPs (Scotter et al., 2006). Type II AFPs from sea raven, smelt and herring is ~125 residues long and cysteine-rich globular proteins containing five disulfide bonds (Ng and Hew, 1992). Type III AFPs are globular proteins of ~65 amino acid residues but with a plane ice-binding surface, and it is isolated from Arctic and Antarctic eel pouts (Sonnichsen et al., 1993). Type IV AFPs are found in longhorn sculpins and they are alpha helical proteins rich in glutamate and glutamine residues (Deng et al., 1997).

The activity of AFPs in over wintering plants was first reported in 1992 (Sidebottom et al., 2000) with low TH activity and high ice RI activity. AFPs prevent plants from the damages caused by cellular dehydration and growth of ice crystals on their surface via the recrystallization inhibition mechanism by inhibiting the formation of extracellular ice. The homologous nature of plant AFPs with PR (β-1, 3-glucanases, chitinases, thaumatin-like protein and polygalacturonase inhibiting protein) protein was evident from their properties of providing a protection to the plants against various psychrophilic pathogens (Davies et al., 2002). The AFPs from plants have been isolated from *Solanum dulcamara, Secale cereale, Daucus carota* and *Lolium perenne* (Kuiper et al., 2001). The antifreeze or ice-recrystallization inhibitory activity is present in overwintering plants only after they have been rendering to low temperatures and only in plants that abide the presence of ice in their tissues. RI activity of AFP has been observed in different parts of over wintering plants such as seeds, stems, flowers, buds, rhizomes, etc. Cold-tolerant plants use a variety of small molecular weight solutes like simple sugars to stabilize the structure of membrane or combat cold-induced osmotic imbalances (Thomashow, 1998). The antifreeze nature of plant AFPs was aided by the presence of characteristic LRR domain in their protein sequences (Meyer et al., 1999).

In the present study, first time, we have developed a sequence and structural based classification scheme for plant AFPs to understand their functions. Some of the classification study was already reported based on extensive phylogenetic analysis (Tyagi et al., 2010). The proposed sequence and structural based classification can be separated into various stages.

At the first stage of the study, various physical and chemical properties of the selected AFPs were computed. At the second stage, we identified the functional domains present in the proteins; however, the functional class of undetermined. AFPs was identified through comparative sequence analysis. At the third stage, the sub-cellular localization and signal peptide cleavage site was predicted, which helped to explore the functional localization of AFPs in the cell. At the fourth stage, we predicted the topology of the proteins and identified the transmembrane regions present in the proteins. At the fifth stage, the N- and O-linked glycosylation sites were predicted, which is essential for antifreeze activity in some of the plant such as *S. dulcamara*. At the sixth stage, we employed comparative sequence analysis to understand the evolutionary relationship of chosen and their related proteins, which can be helped into identifying the functional classes of undetermined AFPs. At the seventh stage, the secondary and 3D structures of AFPs were predicted and compared with reported structures. Finally, the binding site residues and solvation energy was computed, and it can be useful for further molecular interaction studies of AFPs with ice-crystal. Moreover, from an application point of view, the sequence and structural features described here could be used to search for new class of AFPs and their homology in newly sequenced plant genome.

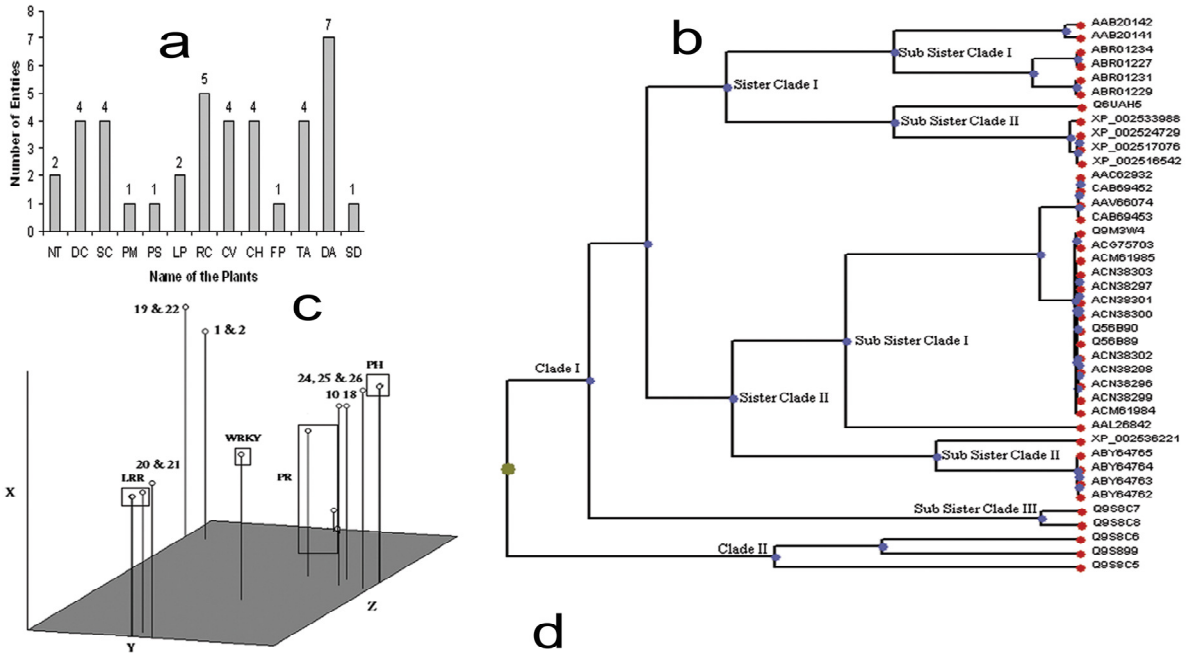
## MATERIALS AND METHODS

### Primary sequence analysis

Antifreeze protein sequences were retrieved from two different databases such as UniProt (www.uniprot.org) (Wu et al., 2006), and GenPept (http://www.ncbi.nlm.nih.gov) (Figure 1a and Table 1). These proteins were subjected into ProtParam web server (www.expasy.org/tools) (Wilkins et al., 1999), for computing various physico chemical properties such as amino acid composition, molecular weight, isoelectric point, instability index, aliphatic index, extinction coefficient and grand average of hydropathicity (GRAVY) score. Functional domains present in the given sequences were identified using SMART web server (www.expasy.org/tools) (Letunic et al., 2009). The sub cellular localization of AFPs was predicted using TargetP (www.expasy.org/tools) and prediction of signal peptides presented in the given sequences were identified through SignalP server (www.expasy.org/tools) (Emanuelsson et al., 2007). The web server SOSUI (www.expasy.org/tools) (Hirokawa et al., 1998) was used for the prediction of transmembrane regions and N and O glycosylation sites of AFPs were predicted using two different servers namely NetNGlyc (http://www.cbs.dtu.dk/services/NetNGlyc/) and NetOGlyc (http://www.cbs.dtu.dk/services/NetOGlyc/).

### Multiple sequence alignment (MSA) and phylogenetic tree construction

Local MSA was performed using Dialign web server (http://bibiserv.techfak.uni-bielefeld.de/dialign/submission.html) (Brudno et al., 2004) with the default threshold value (T = 0), and it finds the region of local similarity. Based on the MSA results, an unrooted phylogenetic tree was constructed using the neighbor-joining (NJ) method. The statistical significance of NJ method was evaluated by a bootstrap analysis with 1000 iterative tree constructions. The “consense” program of Phylip (Retief, 2008) was used to generate a consensus tree, and it was visualized by Phylodraw 0.82 program(Choi et al., 2000). The rectangular binary




d

```
AAB20142          3   AD--------- ---------- ---------- ---------- ----------
AAB20141          2   -D--------- ---------- ---------- ---------- ----------
AAC62932         42   TAL-KNPTIT DS---WVSDD ---DCCGWDL VECDETSNRI ISLIIQDDEA
Q9S8C6            1   ---------- ---------- ---------- ---------- ----------
Q9S8C7            1   ATI-TV-VNK FSYTVW-PGA ---LPFGGVG LG--------- -----PGQ--
Q9S8C5            1   ---------- ---------- ---------- ---------- ----------
Q9S8C8            1   -DsfSI-TNR CAFTVW-PAA ---TPVGG-- ---------- ------GQ--
Q9S899            1   ---------- ---------- ---------- ---------- ----------
Q9M3W4            1   ---------- ---------- ---------- ---------- ----------
Q6UAH5            5   ---------- ---------- ---------- ---------- --------T-
AAV66U74         42   TAL-KNPTIT DS---WVSDD ---DCCGWDL VECDETSNRI ISLIIQDDEA
CAB69452         42   TAL-KNPTIT DS---WVSDD ---DCCGWDL VECDETSNRI ISLIIQDDEA
CAB69453          1   ---------- ---------- ---------- ---------- ----------
XP_002533988      1   ---------- ----------M TMVTP----- -----SEAHP ETMDYLSRT-
XP_002524729      8   -----I-TTR rleqh----S ---PACSMQl ------PESPK VPMDFLSRS-
XP_002517076      2   -DKP-----IT EP---WRPDL ALFRP----- -----PETPR EPMEFLSRS-
XP_002516542     26   SST-KS-LSK KLENIde--N ---GPALWQP VPCDAPETPT ESMEFLARS-
XP_002536221      3   ---------- ---------- ---------- ---------- ----------
ABR01234          4   ---------- ---------- ---------- ---------- ----------
ABR01231          3   GNKP-----IT EQI------- ---------- ---------- ----------
ABR01229          3   GNKP-----IT EQI------- ---------- ---------- ----------
ABR01227          4   ---------- ---------- ---------- ---------- ----------
ABY64765         22   ---------- ---------- ---------- ---------- ----------
ABY64764         26   ---------- ---------- ---------- ---------- ----------
ABY64763         23   ---------- ---------- ---------- ---------- ----------
ABY64762         20   ---------- ---------- ---------- ---------- ----------
ACG75703          1   ---------- ---------- ---------- ---------- ----------
ACM61985          1   ---------- ---------- ---------- ---------- ----------
ACM61984         26   GNL-SG-GGV LLRAMW-SGA ---SCCGWEG VGCDGQSGRV MALLLPGRG-
ACN38303         37   GNL-SN-GGV LLHAKW-PDN ---PCCSWEG VGCDGGSGRV TTLWLPGHG-
ACN38302         37   RNL-GGVEGV LVRAAW-FGA ---SCCSWEG VGCDDATGRV TVLQFPTRG-
ACN38301         38   DNL-KG-GGV LLHAAW-SGA ---SCCGWEG VGCDDASGRV TTLWLPRCG-
ACN38300         36   DNL-TG-GGV LLRAAW-SGT ---SCCGWEG VGCDGTSGRV TTLWLPRRG-
ACN38299         36   GNL-SG-GGV LLRSVW-SGD ---SCCGWEG VGCDDASGRV TTMWLPRRG-
ACN38298         37   GNL-GSPGGV LLRAAW-SGA ---SCCDWEG VSCDGASGRV TALQLPTRG-
ACN38297         37   GNL-SG-GAV LLRATW-SGA ---SCCGWEG VGCDGASGHV TTLWLPGRG-
ACN38296         37   RNL-GGVGGV LLRAAW-SGD ---GCCDWEG VGCDGASGRV TTLQLPTRG-
Q56B90           36   GNL-SG-GAA LLRAAW-SGA ---SCCVWEG VNCDGTSGRV TALRLPGHG-
Q56B89           35   GTL-GA-VAR HLHSVW-SGA ---SCCDWEG VGCHGASGRV TALRLHGHG-
AAL26842         35   TTS-KS-ISG ELQKr------ ----------L SQCDKGINms qsnqea----

                      ********** ********** ********** ********** **********
                      ***  **  *** ****** *** ******* ********** *********
                      ***  **  *** ****** *** ******* ********** *********
                                   *      *  ******* ********** *****
```

**Figure 1.** Comparative sequence analysis of antifreeze proteins in over wintering plants. **(a)** Bar diagram showing number of entries present in the protein sequence databases (At the time of study). Species abbreviations followed in the diagram: NT, *Nicotiana tabacum,* DC, *D. carota*, SC, *S. cereale*, PM, *P. monticola*, PS, *P. suaveolens*, LP, *L. perenne*, RC, *R. communis*, CV, *C. vulgaris*, CH, *Chlamydomonas sp. CCMP681*, FP, *F. pratensis*, TA, *T. aestivum*, DA, *D. antarctica*, SD, *S. dulcamara*. **(b)** NJ tree of antifreeze proteins was constructed by importing the numeric output of Dialign web server into Phylip - consense program to generate a consensus tree and it was visualized by Phylodraw Version 0.8. **(c)** The Principal coordinate analysis of antifreeze proteins was performed by NTSYS pc. Expansion of codes in this 3D plot is: 1 & 2 = *N. tabacum*, LRR = Leucine Rich Repeat (*D. carota*, *F. pratensis*, *T. aestivum, L. perenne* and *D. antarctica*), PH = Pleckstrin Homology (*R. communis*), PR = Pathogenesis Related (*Secale cereale* & *Pinus monticola*), WRKY = DNA binding protein from *S. dulcamara*, 10 & 18 = *P. suaveolens* and *R. communis*, 19, 22 = *C. vulgaris* and 24, 25 and 26 = *Chlamydomonas sp. CCMP681*. **(d)** A portion of multiple sequence alignment of antifreeze proteins was performed using Dialign web server. "*" indicate the degree of local similarity among the sequences. Residues in the blocks represent sequentially conserved regions (SCRs). Lower-case letters denote residues not belonging to any of these selected diagonals or segment pairs and which are not considered to be aligned by Dialign.

**Table 1.** List of plant antifreeze protein sequences retrieved from GenPept and UniProt database.

| S/No. | Accession number | Protein name | Source | Database |
|---|---|---|---|---|
| 1 | AAB20142 | Afa5, antifreeze protein | *Nicotiana tabacum* | GenPept |
| 2 | AAB20141 | Afa3-antifreeze protein | *N. tabacum* | GenPept |
| 3 | AAC62932 | Antifreeze protein | *Daucus carota* | GenPept |
| 4 | Q9S8C6 | 32 kDa antifreeze protein | *Secale cereale* | UniProt |
| 5 | Q9S8C7 | 25 kDa antifreeze protein | *S. cereale* | UniProt |
| 6 | Q9S8C5 | 35 kDa antifreeze protein | *S.e cereale* | UniProt |
| 7 | Q9S8C8 | 16 kDa antifreeze protein | *S.e cereale* | UniProt |
| 8 | Q9S899 | 18.4 kDa candidate antifreeze protein | *Pinus monticola (Western white pine)* | UniProt |
| 9 | Q9M3W4 | Ice recrystallisation inhibition protein | *Lolium perenne (Perennial ryegrass)* | UniProt |
| 10 | Q6UAH5 | Antifreeze protein | *Populus suaveolens* | UniProt |
| 11 | AAV66074 | Antifreeze protein | *Daucus carota* | GenPept |
| 12 | CAB69452 | Antifreeze protein | *D. carota* | GenPept |
| 13 | CAB69453 | Antifreeze protein | *D. carota* | GenPept |
| 14 | XP_002533988 | Putative ice-binding protein | *Ricinus communis* | GenPept |
| 15 | XP_002524729 | Putative ice-binding protein | *R.communis* | GenPept |
| 16 | XP_002517076 | Putative ice-binding protein | *R. communis* | GenPept |
| 17 | XP_002516542 | Putative ice-binding protein | *R. communis* | GenPept |
| 18 | XP_002536221 | Putative ice-binding protein | *R. communis* | GenPept |
| 19 | ABR01234 | Antifreeze protein | *Chlorella vulgaris* | GenPept |
| 20 | ABR01231 | Antifreeze protein | *C. vulgaris* | GenPept |
| 21 | ABR01229 | Antifreeze protein | *C. vulgaris* | GenPept |
| 22 | ABR01227 | Antifreeze protein | *C. vulgaris* | GenPept |
| 23 | ABY64765 | Ice-binding protein-4 | *Chlamydomonas sp. CCMP681* | GenPept |
| 24 | ABY64764 | Ice-binding protein-3 | *Chlamydomonas sp. CCMP681* | GenPept |
| 25 | ABY64763 | Ice-binding protein-2 | *Chlamydomonas sp. CCMP681* | GenPept |
| 26 | ABY64762 | Ice-binding protein-1 | *Chlamydomonas sp. CCMP681* | GenPept |
| 27 | ACG75703 | Ice recrystallization inhibition protein | *Festuca pratensis* | GenPept |
| 28 | ACM61985 | Ice recrystallization inhibition protein 4 | *Triticum aestivum* | GenPept |
| 29 | ACM61984 | Ice recrystallization inhibition protein 3 | *T. aestivum* | GenPept |
| 30 | ACN38303 | Ice recrystallization inhibition protein 1 | *Lolium perenne* | GenPept |
| 31 | ACN38302 | Ice recrystallization inhibition protein 7 | *Deschampsia antarctica* | GenPept |
| 32 | ACN38301 | Ice recrystallization inhibition protein 6 | *D. antarctica* | GenPept |
| 33 | ACN38300 | Ice recrystallization inhibition protein 5 | *D. antarctica* | GenPept |
| 34 | ACN38299 | Ice recrystallization inhibition protein 4 | *D. antarctica* | GenPept |
| 35 | ACN38298 | Ice recrystallization inhibition protein 3 | *D. antarctica* | GenPept |
| 36 | ACN38297 | Ice recrystallization inhibition protein 2 | *D. antarctica* | GenPept |
| 37 | ACN38296 | Ice recrystallization inhibition protein 1 | *D. antarctica* | GenPept |
| 38 | Q56B90 | Ice recrystallization inhibition protein 1 | *Triticum aestivum* | UniProt |
| 39 | Q56B89 | Ice recrystallization inhibition protein 2 | *T.aestivum* | UniProt |
| 40 | AAL26842 | Thermal hysteresis protein STHP-64 | *Solanum dulcamara* | UniProt |

data matrix was created, and all the data analysis was performed by Numerical Taxonomy System, NTSYS-pc ver. 2.02 (Applied Biostatistic, Exeter Software, Setauket, New York, USA). The PCoA analysis of the pair wise genetic distances was also conducted for validating the results of phylogenetic analysis using NTSYS-pc package.

**Secondary and three-dimensional structure predictions**

Out of 40 AFPs, only 13 proteins have their own structural details. Therefore, the secondary structure of remaining 27 plant AFPs was predicted using GOR web server (www.expasy.org/tools) (Sen et al., 2005). In Homology Modeling, the template sequences were

selected from the Protein BLAST search (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) (Altschul et al., 1997) with the help of Protein Data Bank (http://www.rcsb.org). The 3D structure for 13 AFPs showing reasonable sequence similarity with their templates, and their structures were modeled through homology modeling using Modeller 9V4 (Eswar et al., 2008), whereas structures for remaining 14 AFPs were modeled through threading approach using Phyre server (http://www.sbg.bio.ic.ac.uk/~phyre/) (Kelley and Sternberg, 2009). Gromacs 3.3.1(Hess, 2008) was used to refine the 3D models and Molecular dynamic calculations for the generated models soaked in the triclinic single point charge (SPC) water molecule system were carried out for 10 ps and 5000 steps were employed for the calculations. The quality of the generated models was investigated using RamPage (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) (Lovell et al., 2003) and combinatorial extension (CE) (http://cl.sdsc.edu/ce.html) (Shindyalov and Bourne, 1998). Salvation energy/solvent accessible surface area was predicted through Get Area web server (http://curie.utmb.edu/getarea.html) (Fraczkiewics and Braun, 1998). The binding site analysis of generated protein models were performed using theoretical macroscopic titration curves (Thematics) web server (http://pfweb.chem.neu.edu/thematics/submit.html) (Ondrechen et al., 2001).

## RESULTS AND DISCUSSION

### Primary sequence analysis

#### *Physicochemical properties*

Calculation of physicochemical properties by conventional *invitro* methods besides being expansive, is time consuming and cumbersome. The *insilico* physicochemical property prediction may be enhancing our knowledge for experimental design. The amino acid sequence of plant AFPs revealed that eighteen of the plant AFPs was hydrophobic; twenty one were hydrophilic in nature due to the presence of polar and non-polar amino acid residues in their protein sequence. The AFP from *S. cereale* (Q9S8C5) has an equal number of polar and non-polar residues. The calculated isoelectric point (pI) value of 21 AFPs was less than 7, and it indicated that these were acidic and remaining 19 proteins were basic in nature. This result will be essential for developing buffer systems for purification of proteins by isoelectric focusing and two-dimensional electrophoresis. Extinction coefficient (EC) value calculated for the given entries revealed that the following four proteins ACG75703, AAB20141, Q9M3W4 and AAB20142 cannot be studied by UV spectral method. Instability index calculations identified twenty seven entries as stable proteins with the values were smaller than 40 and remaining proteins may be unstable. Most of the entries were found to have a good aliphatic index score which indicated that these AFPs were thermodynamically stable over a wide range of temperatures. Calculated GRAVY index values are normally used as a guide to help, predict, with a great measure of uncertainty, the probability that the protein will produce crystals and therefore, be amendable for X-ray crystallographic analysis. In our study, twenty seven proteins were showing good GRAVY index value where as others were found to have optimum values. The primary sequence analysis of the selected proteins was listed in Tables 2 and 3, respectively.

#### *Prediction of signal peptide cleavage site and sub cellular localization*

The regions 14 to 15 of ACM61985, 15 to 16 of ACM61984, 18 of AAB20141, 20 to 21 of ACN38300, ACN38299, Q56B90, Q56B89, 21 to 22 of ACN38303, ACN38302, ACN38298, ACN38297, ACN38296, 22 to 23 of ACN38301, 23 to 24 of ABY64762, 25 to 26 of CAB69453, ABY64765, 26 to 27 of AAC62932, AAV66074, CAB69452, ABY64763, 29 to 30 of ABY64764 and 31 of AAB20142 indicated the presence of N-terminal signal peptides, which implies that these proteins could be targeted through the secretory pathway. The AFP from *N. tabacum, D. carota, T. aestivum, D. antarctica, L. perenne* (ACN38303) *and Chlamydomonas sp. CCMP681* contained putative signal peptide sequences, suggesting that these AFPs were secreted and primarily function in extra cellular space. However, *S. cereale, P. monticola, L. perenne* (Q9M3W4), *R. communis, C. vulgaris, F. pratensis* and *S. dulcamara* have a type of AFP that did not contain putative signal peptide sequence, and it is therefore, likely to remain intracellular.

#### *Analysis of domain*

Apart from the signal peptide, sixteen of the plant AFPs has a LRR (and LRRNT), four AFPs have PH and one AFP has WRKY domain. The sequence of *D. carota* - AFP is similar to that of PGIPs and contained LRRs, which exhibit antifreeze activity (Meyer et al., 1999). Similarly, the AFP from *L. perenne, T. aestivum* and *D. antarctica* is also containing LRR domain, which also showed the antifreeze activity. The functional class of undetermined AFPs was identified through comparative sequence analysis.

#### *Analysis of glycosylation sites and transmembrane regions*

By using the NetNGlyc and NetOGlyc web servers, the amino acid sequence of AFPs was examined for possible N- and O-linked glycosylation. In our study, the AFPs (Q9S8C6 and Q6UAH5) from *S. cereale* and *P. suaveolens* did not contain glycosylation sites, where as remaining AFPs were found to have several glycosylation sites (N-Linked: 24, O-Linked: 34 and Both N and O Linked: 20). Glycosylation, a post translational modification,

**Table 2.** Various physicochemical parameters of plant antifreeze proteins computed using ProtParam web server.

| S/No. | Accession number | Molecular weight | Isoelectric point | Instability index | Aliphatic index | Extinction coefficient | Grand average of hydropathicity (GRAVY) |
|---|---|---|---|---|---|---|---|
| 1 | AAB20142 | 5109.6 | 3.56 | 8.04 | 96.13 | - | 1.134 |
| 2 | AAB20141 | 3359.6 | 4.43 | 1.131 | 81.05 | - | 0.563 |
| 3 | AAC62932 | 36742.1 | 5.02 | 42.12 | 99.00 | 23210 | -0.064 |
| 4 | Q9S8C6 | 2093.3 | 5.83 | 51.87 | 116.50 | 1490 | 0.395 |
| 5 | Q9S8C7 | 2934.3 | 8.63 | 26.58 | 87.24 | 6990 | 0.514 |
| 6 | Q9S8C5 | 1549.6 | 4.00 | 43.18 | 36.88 | 125 | -0.588 |
| 7 | Q9S8C8 | 2482.7 | 5.83 | 34.14 | 52.92 | 5500 | 0.100 |
| 8 | Q9S899 | 2419.6 | 4.79 | 77.00 | 79.55 | 5500 | -0.368 |
| 9 | Q9M3W4 | 11766.1 | 5.17 | 11.54 | 61.61 | - | -0.679 |
| 10 | Q6UAH5 | 17548.1 | 10.02 | 55.42 | 67.15 | 19605 | -0.550 |
| 11 | AAV66074 | 36825.1 | 4.87 | 43.21 | 97.83 | 23210 | -0.087 |
| 12 | CAB69452 | 36845.3 | 4.99 | 43.35 | 98.70 | 23210 | -0.056 |
| 13 | CAB69453 | 21785.9 | 5.78 | 37.91 | 93.55 | 4720 | -0.104 |
| 14 | XP_002533988 | 39792.6 | 5.87 | 31.19 | 82.19 | 43930 | -0.185 |
| 15 | XP_002524729 | 51873.7 | 9.22 | 46.69 | 78.17 | 34085 | -0.298 |
| 16 | XP_002517076 | 51022.9 | 9.27 | 41.75 | 81.66 | 44710 | -0.286 |
| 17 | XP_002516542 | 45915.0 | 9.06 | 44.94 | 77.14 | 47940 | -0.423 |
| 18 | XP_002536221 | 30830.4 | 4.58 | 3.45 | 82.65 | 8940 | 0.247 |
| 19 | ABR01234 | 18682.6 | 8.68 | 21.27 | 62.81 | 1490 | -0.810 |
| 20 | ABR01231 | 10803.7 | 6.07 | 21.99 | 52.88 | 5500 | -0.903 |
| 21 | ABR01229 | 10775.7 | 6.07 | 23.44 | 51.06 | 5500 | -0.926 |
| 22 | ABR01227 | 18710.6 | 8.68 | 21.11 | 63.88 | 1490 | -0.796 |
| 23 | ABY64765 | 36814.1 | 4.48 | 32.85 | 73.58 | 28960 | 0.147 |
| 24 | ABY64764 | 36975.7 | 4.66 | 33.12 | 85.40 | 42815 | 0.254 |
| 25 | ABY64763 | 36507.0 | 4.48 | 34.32 | 81.69 | 41325 | 0.209 |
| 26 | ABY64762 | 36523.8 | 4.46 | 41.36 | 69.72 | 35950 | 0.011 |
| 27 | ACG75703 | 11555.1 | 7.54 | 11.62 | 65.74 | - | -0.492 |
| 28 | ACM61985 | 18245.8 | 9.23 | 20.05 | 78.29 | 8480 | -0.491 |
| 29 | ACM61984 | 19046.2 | 9.65 | 41.16 | 106.74 | 18240 | 0.188 |
| 30 | ACN38303 | 29135.7 | 9.04 | 25.83 | 93.51 | 30855 | -0.070 |
| 31 | ACN38302 | 22803.1 | 9.04 | 29.49 | 79.37 | 16750 | -0.192 |
| 32 | ACN38301 | 29252.9 | 9.12 | 38.14 | 93.24 | 27875 | -0.040 |
| 33 | ACN38300 | 29195.7 | 8.56 | 35.31 | 89.32 | 23865 | -0.071 |
| 34 | ACN38299 | 23021.3 | 5.72 | 32.89 | 82.06 | 18490 | -0.166 |

**Table 2.** Cont'd

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 35 | ACN38298 | 22011.1 | 9.01 | 30.78 | 78.99 | 18240 | -0.162 |
| 36 | ACN38297 | 30514.0 | 9.58 | 36.94 | 89.80 | 30855 | -0.158 |
| 37 | ACN38296 | 22560.0 | 9.86 | 27.25 | 83.38 | 11250 | -0.125 |
| 38 | Q56B90 | 29104.1 | 8.23 | 25.37 | 99.96 | 27305 | 0.114 |
| 39 | Q56B89 | 42886.5 | 8.60 | 29.46 | 102.69 | 26845 | 0.051 |
| 40 | AAL26842 | 64757.9 | 6.27 | 52.77 | 63.11 | 29255 | -0.969 |

**Table 3.** Primary sequence analysis of plant antifreeze proteins was predicted by using various web servers [Domain: SMART, signal peptide: Signal P, sub cellular localization: Target P, transmembrane region: SOSUI and glycosylation sites: NetNGlyc and NetOGlyc] and expansion of domains [LRR: Leucine rich repeat, PH: Pleckstrin homology, ND: Not determined, *: Domain are identified through comparative sequence analysis].

| Accession number | Domain | | Signal Peptide | Sub cellular | Transmembrane region | | Glycosylation | |
|---|---|---|---|---|---|---|---|---|
| | Name | Position | cleavage site position | localization | Number | position | N-linked | O-linked |
| AAB20142 | ND | ND | 31 | Secretory pathway | No | No | No | Yes |
| AAB20141 | ND | ND | 18 | Secretory pathway | No | No | No | Yes |
| AAC62932 | LRR-1, LRRNT-2 | 30-68, 266-288 | 26,27 | Secretory pathway | 1 | 4-26 | Yes | No |
| Q9S8C6 | ND | ND | No | Other | No | No | No | No |
| Q9S8C7 | ND | ND | No | Other | No | No | No | Yes |
| Q9S8C5 | ND | ND | No | Other | No | No | No | Yes |
| Q9S8C8 | ND | ND | No | Other | No | No | No | Yes |
| Q9S899 | ND | ND | No | Other | No | No | No | Yes |
| Q9M3W4 | ND | ND | No | Chloroplast | No | No | No | Yes |
| Q6UAH5 | PH* | ND | No | Mitochondrion | No | No | No | No |
| AAV66074 | LRRNT-2<br>LRR-1<br>LRR-1<br>LRR-1 | 30-68<br>147-169<br>171-193<br>266-268 | 26,27 | Secretory pathway | 1 | 4-26 | Yes | No |
| CAB69452 | LRRNT-2<br>LRR-1<br>LRR-1<br>LRR-1 | 30-68<br>147-169<br>171-193<br>266-268 | 26,27 | Secretory pathway | 1 | 4-26 | Yes | No |

**Table 3.** Cont'd

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| CAB69453 | LRR-1 | 131-153 | 25,26 | Secretory pathway | No | No | Yes | No |
| XP_002533988 | PH-2 | 244-350 | No | Other | No | No | Yes | Yes |
| XP_002524729 | PH-2 | 364-474 | No | Other | No | No | Yes | Yes |
| XP_002517076 | PH-2 | 362-472 | No | Other | No | No | Yes | Yes |
| XP_002516542 | PH-2 | 297-406 | No | Chloroplast | No | No | Yes | Yes |
| XP_002536221 | ND | ND | No | Other | No | No | Yes | Yes |
| ABR01234 | ND | ND | No | Chloroplast | No | No | No | Yes |
| ABR01231 | ND | ND | No | Other | No | No | No | Yes |
| ABR01229 | ND | ND | No | Other | No | No | No | Yes |
| ABR01227 | ND | ND | No | Chloroplast | No | No | No | Yes |
| ABY64765 | ND | ND | 25,26 | Secretory pathway | No | No | Yes | Yes |
| ABY64764 | ND | ND | 29,30 | Secretory pathway | 2 | 15-37, 43-65 | Yes | Yes |
| ABY64763 | ND | ND | 26,27 | Secretory pathway | 2 | 11-33, 41-63 | Yes | Yes |
| ABY64762 | ND | ND | 23,24 | Secretory pathway | No | No | Yes | Yes |
| ACG75703 | LRR* | ND | No | Chloroplast | No | No | Yes | Yes |
| ACM61985 | LRR-1 | 13-35 | 14,15 | Secretory pathway | No | No | Yes | Yes |
| ACM61984 | LRRNT-2<br>LRR-1 | 14-53<br>83-105 | 15,16 | Secretory pathway | No | No | Yes | Yes |
| ACN38303 | LRRNT-2<br>LRR-1 | 25-64<br>65-93 | 21,22 | Secretory pathway | 1 | 8-28 | Yes | Yes |
| ACN38302 | LRRNT-2 | 25-65 | 21,22 | Secretory pathway | 2 | 6-27, 40-62 | No | Yes |
| ACN38301 | LRRNT-2<br>LRR-1 | 25-65<br>95-117 | 22,23 | Secretory pathway | 3 | 2-23, 45-67, 75-97 | No | Yes |
| ACN38300 | LRRNT-2<br>LRR-1 | 24-63<br>93-115 | 20,21 | Secretory pathway | 2 | 1-22, 39-60 | Yes | Yes |
| ACN38299 | LRRNT-2 | 24-63 | 20,21 | Secretory pathway | No | No | Yes | Yes |
| ACN38298 | LRRNT-2 | 25-65 | 21,22 | Secretory pathway | 1 | 5-26 | No | Yes |
| ACN38297 | LRRNT-2 | 25-64 | 21,22 | Secretory pathway | 2 | 5-26, 41-63 | Yes | Yes |
| ACN38296 | LRRNT-2 | 25-65 | 21,22 | Secretory pathway | No | No | Yes | Yes |

**Table 3.** Cont'd

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Q56B90 | LRRNT-2<br>LRR-1<br>LRR-1 | 24-63<br>93-115<br>117-139 | 20,21 | secretory pathway | 3 | 2-23, 38-60, 75-97 | Yes | Yes |
| Q56B89 | LRR-1<br>LRR-1<br>LRR-1<br>LRR-1<br>LRR-1<br>LRR-1 | 90-114<br>138-162<br>187-211<br>212-235<br>236-260<br>262-284 | 20,21 | Secretory pathway | 1 | 1-22 | Yes | Yes |
| AAL26842 | WRKY | 182-250 | No | other | No | No | Yes | Yes |

is not essential for antifreeze activity in some cases such as *D. carota* (Worrall et al., 1998). However, the AFP from *S. dulcamara* lost the activity after the removal of their glycan moiety (Huang and Duman, 2002). Therefore, the analysis of glycosylation sites is also one of the important parameter to understand the mechanism of ice-recrystallization inhibition.

From the results of SOSUI web server, it was concluded that the thirteen of the given entries were classified as membrane proteins and the remaining were classified as soluble proteins.

## Comparative sequence analysis

In this study, we have used comparative sequence analysis approach to categorize the AFPs based on sequence level similarity, identity and pair wise distances, which provide information about the homology of particular AFP. It includes multiple sequence alignment followed by phylogenetic analysis. The PCoA was used to validate the results for phylogenetic analysis of all AFPs.

### *Multiple sequence alignment, phylogenetic analysis and principal coordinate analysis*

Dialign web server constructed the local alignment from the gap free pairs of segments of the protein sequences. The results for MSA (Figure 1d) of plant AFPs explore that, there was a considerable variation among the proteins (Number of variable sites: 804), with extensive variation in both amino and carboxy terminals. Remarkably, 198 conserved sites were observed in the amino acid residues of the plant AFPs and they were interspersed throughout the alignment. The overall mean distance of fourty AFPs from various plants was 1.589. In order to determine the evolutionary relationship of these AFPs, the phylogenetic tree was constructed from various plants such as *N. tabacum*, *D. carota*, *S. cereale*, *P. monticola*, *L. perenne*, *P. suaveolens*, *R. communis*, *C. vulgaris*, *Chlamydomonas sp. CCMP681*, *F. pratensis*, *T. aestivum*, *D. antarctica* and *S. dulcamara*. NJ phylogenetic analysis of the plant AFP sequences resulted in the clustering of two major clades (Figure 1b) and their details were given as follows.

**Clade I:** This is strongly supported group, and it composed of numerous sister cades (SC) such as SC1, SC2 and SC3. The SC1 has two sub sister clades (SSC) namely SSC1 [AAB20142, AAB20141 (*N. tabacum*), ABR01234, ABR01227, ABR01231, ABR01229 (*C. vulgaris*)] and SSC2 [Q6UAH5 (*P. suaveolens*) and PH domain: XP_002533988, XP_002524729, XP_002517076, XP_002516542 (*R. communis*)]. SC2 also has two SSCs and they are SSC1 [AAC62932, AAV66074, CAB69452, CAB69453 (*D. carota*), Q9M3W4 (*L. perenne*), ACG75703 (*F. pratensis*), ACM61985 (*T. aestivum*), ACN38303 (*L. perenne*), ACN38297, ACN38301, ACN38300 (*D. Antarctica*), Q56B90, Q56B89 (*T. aestivum*), ACN38302, ACN38298, ACN38296, ACN38299 (*D. antarctica*), ACM61984 (*T. aestivum*), AAL26842 (*S. dulcamara*)], in which eighteen AFPs are belonging to LRR domain and a AFP belongs to WRKY DNA binding domain and SSC2 [XP_002536221 (*R. communis*), ABY64765, ABY64764, ABY64763, ABY64762

(*Chlamydomonas sp. CCMP681*)]. In addition, the SC3 contained PR related AFPs such as Q9S8C7 and Q9S8C8 (*S. cereale*). Homogenous (Same taxas: SC3) andheterogeneous (Different taxas: SC1 and SC2 with its Sub Clades) sister clades were obtained by the aforementioned observation. In addition to this, five mono taxas were also observed (Q6UAH5, XP_002533988, ACM61985, AAL26842, XP_002536221), which are closely related to SC1 and SC2. Based on above observation, the AFP from *F. pratensis* is newly identified LRR domain containing protein, which is closely related to AFP of *D. antarctica, D. carota, T. aestivum* and *L. perenne*. The AFP from *P. suaveolens* (Q6UAH5) exists as mono taxa, and it is closely related to PH domain containing proteins, which revealed that this protein mighthave PH like activity. The AFP – domains of *C. vulgaris* (ABR01231, ABR01229, ABR01227, and ABR01234) and *R. communis* (XP_002536221) are undetermined. More detailed comparative sequence analysis was performed on the above mentioned proteins. The AFPs from *N. tabacum* resemble type III *Winter flounder* AFP. One of the undetermined AFP (XP_002536221) from *R. communis* is similar with hemagglutinin-related protein from *Granulibacter bethesdensis CGDNIH1* (YP_744956 and YP_745229) and remaining proteins are belonging to PH family. *C. vulgaris* AFPs are similar with some of the hypothetical proteins from *Desulfovibrio piger ATCC 29098* (ZP_03311955), *Rhodobacter sphaeroides 2.4.1* (YP_352443), *Rhodobacter sphaeroides KD131* (YP_02525057). Similarity, the AFPs from *Chlamydomonas sp. CCMP681* showed similarity with some of the hypothetical proteins from bacteria (Raymond et al., 2009).

**Clade II:** This is a very small clade compared to clade I, which comprised of Q9S8C6 (*S. cereale*), Q9S899 (*P. monticola*) and Q9S8C5 (*S. cereale*). It has no further sub divisions. *S. cereale* AFPs are homologous to PR proteins, which also found in clade II (Q9S8C6 and Q9S8C5), and it has both antifreeze as well as enzymatic activities. It provides a protection against psychrophilic pathogens. The AFPs from *S. cereale* is similar to the member of three classes of PR related proteins such as endochitinases, endo-β-1, 3 glucanases and thaumatin like proteins (Hon et al., 1995). The domain sequence of Q9S899 (*P. monticola*) was undetermined and however, it was found to be closely related to *S. cereale* - AFP. The pair wise distance between *S. cereale* and *P. monticola* was 0.673 (sequence identity: 57.1% and sequence similarity: 71.4%), which was very low in comparison with others indicating that less divergence has occurred. Based on the aforementioned observation, it was concluded that, the *P. monticola* – AFP is a newly identified PR family protein. All the AFPs were well supported in bootstrap analysis (100%).

The 3D plot (Figure 1c) of PCoA is a diverse combination of various AFP data. Several closely related plant groups were observed through this three-dimensional plot such as LRR (*D. carota:* AAC62932, AAV66074, CAB69452, CAB69453, *L. perenne:* Q9M3W4, ACN38303, *T. aestivum:* ACM61985, ACM61984, Q56B90, Q56B89, *D. antarctica:* ACN38302, ACN38301, ACN38300, ACN38299, ACN38298, ACN38297, ACN38296 and *F. pratensis*: ACG75703), PH (*R. communis:* XP_002533988, XP_002524729, XP_002517076 and XP_002516542) and PR (*S. cereale:* Q9S8C6, Q9S8C7, Q9S8C5, Q9S8C8 and *P. monticola:* Q9S899). Both phylogenetic and PCoA analysis revealed that, the newly identified *F. pratensis* and *P. monticola* – AFPs are belonging to LRR and PR family, respectively. WRKY domain containing AFP from *S. dulcamara* was deviated significantly from other Plant AFPs, which is located in the central position of Mod3D plot. In addition, some of the plant - AFPs are existing as mono taxa as phylogenetic analysis such as *P. monticola, P. suaveolens, R. communis*, etc. Hence, PcoA agreed well with the results of phylogenetic analysis.

### *Secondary structure prediction*

Secondary structural study is significant as it provides direct imminent into the functional role of a protein, and it can be a preliminary step in the direction towards the prediction of 3D structures by fold recognition or threading. Secondary and 3D structure of AFPs from *L. perenne, (*Kuiper et al., 2001*) D. antarctica (*John et al., 2009*) and D. carota (*Zhang et al., 2004*)* were reported previously. Various secondary structural classes (alpha: 8, beta: 11 and alpha + beta: 8) of AFPs were observed from the results of ProClass web server (Raghava, 1999), in which, most of the LRR domain containing proteins are exist as Alpha + Beta class. Alpha helices are dominated in some of the AFPs such as *N. tabacum* and *C. vulgaris*. The secondary structure of *S. cereal*e AFPs looks coil like structure, which is similar to the PR proteins and posses, both antifreeze as well as antifungal activity. The composition of secondary structural elements from the various AFPs was represented in Table 4. The secondary structure of four AFPs (XP_002533988, XP_002524729, XP_002517076 and XP_002516542) from *R. communis* belongs to beta class, which contained PH domain, and it may be essential for intracellular signaling. A newly identified LRR domain containing AFP from *F. pratensis* contain only beta pleated sheets, which is closely related to AFP of *D. antarctica, D. carota, T. aestivum* and *L. perenne*. The secondary structure composition of this protein is Ee: 38.26% and Cc: 61.74%, which is more similar to the reported AFP structure (Kuiper et al., 2001) from *L. perenne* (E: 36.44% and C: 63.56%). The secondary structural class of AFP from R. *communis* (XP_002536221) is beta as hemagglutinin-related proteins from several bacteria. The coil composition of

**Table 4.** Secondary structural classes and their composition of plant antifreeze proteins.

| Accession number | Secondary structural class | Secondary structure composition (%) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Hh | Gg | Ii | Bb | Ee | Tt | Ss | Cc |
| AAB20142 | Alpha | 90.32 | - | - | - | - | - | - | 9.68 |
| AAB20141 | Alpha | 84.21 | - | - | - | - | - | - | 15.79 |
| Q9S8C6 | Beta | - | - | - | - | 35 | - | - | 65 |
| Q9S8C7 | Beta | - | - | - | - | 48.28 | - | - | 51.72 |
| Q9S8C5 | Beta | - | - | - | - | 12.50 | - | - | 57.50 |
| Q9S8C8 | Beta | - | - | - | - | 33.33 | - | - | 66.67 |
| Q9S899 | Alpha | 50 | - | - | - | 9.09 | - | - | 40.91 |
| Q6UAH5 | Alpha + Beta | 37.75 | - | - | - | 19.87 | - | - | 42.38 |
| XP_002533988 | Beta | 24.17 | - | - | - | 33.33 | - | - | 42.50 |
| XP_002524729 | Beta | 13.37 | - | - | - | 38.48 | - | - | 48.15 |
| XP_002517076 | Beta | 10.90 | - | - | - | 43.61 | - | - | 45.49 |
| XP_002516542 | Beta | 8.29 | - | - | - | 52.37 | - | - | 39.34 |
| XP_002536221 | Beta | 1.81 | - | - | - | 39.76 | - | - | 58.43 |
| ABR01234 | Alpha | 74.16 | - | - | - | 4.49 | - | - | 21.35 |
| ABR01231 | Alpha | 68.27 | - | - | - | 4.81 | - | - | 26.92 |
| ABR01229 | Alpha | 70.19 | - | - | - | 4.81 | - | - | 25 |
| ABR01227 | Alpha | 71.35 | - | - | - | 4.49 | - | - | 24.16 |
| ABY64765 | Beta | 9.22 | - | - | - | 31.28 | - | - | 59.5 |
| ABY64764 | Beta | 13.09 | - | - | - | 25.07 | - | - | 61.84 |
| ABY64763 | Beta | 10.96 | - | - | - | 27.25 | - | - | 61.80 |
| ABY64762 | Beta | 7.93 | - | - | - | 30.31 | - | - | 61.76 |
| ACG75703 | Beta | - | - | - | - | 38.26 | - | - | 61.74 |
| ACM61985 | Beta | 13.14 | - | - | - | 25.14 | - | - | 61.71 |
| ACM61984 | Alpha+Beta | 29.28 | - | - | - | 19.89 | - | - | 50.83 |
| Q56B90 | Alpha+Beta | 18.93 | - | - | - | 28.27 | - | - | 52.86 |
| Q56B89 | Alpha+Beta | 30.32 | - | - | - | 19.56 | - | - | 50.12 |
| AAL26842 | Alpha+Beta | 18.78 | - | - | - | 19.80 | - | - | 61.42 |

Hh: Alpha helix, Gg: $3_{10}$ helix, Ii: Pi helix, Bb: Beta bridge, Ee: Extended strand, Tt: Beta turn, Ss: Bend region, Cc: Coil.

AFP from *P. suaveolens* is nearly similar to PH domain containing AFP of *R. communis*. The predicted secondary structure class of *C. vulgaris* and *Chlamydomonas sp. CCMP681* AFPs is alpha and beta, which is generally consistent with their predicted models. The secondary structure elements of *S. dulcamara* AFP contained five beta pleated sheets, which is similar to the experimental structure of *A. thaliana* WRKY DNA binding domain (Yamasaki et al., 2005) and it belongs to beta class.

## Three-dimensional structure prediction and binding site analysis

In the result of $BLAST_P$ search, only 13 AFPs were showing reasonable identity (> = 40%) with their templates. So, the 3D structures for these proteins were predicted using Modeller 9V4 based on homology modeling approach. A threading method was also adopted for predicting the 3D structure of the remaining 14 AFPs using Phyre server, because $BLAST_P$ provided the low sequence identity (Less than 40%) structural homologs or templates. To eliminate the distortion in geometry, the predicted homology models (Figure 2) were refined by consecutive iterations of MD simulation followed by energy minimization using Gromacs 3.3.1. MD simulation was carried out on all the predicted models in an aqueous environment (SPC water molecules). During the MD simulation process, the root mean square deviation (RMSD) of the predicted models and their corresponding templates were studied, which are less than 2. MD simulation results indicated that our predicted models were more stable and they will make good interaction with ice crystal. In this step, the quality of the predicted models was also improved. The optimized models were subjected into internal assessment of self-consistency checks such as stereo chemical quality to locate the divergences from normal bond lengths, dihedrals and non-bonded atom-atom distances. No spurious angle to

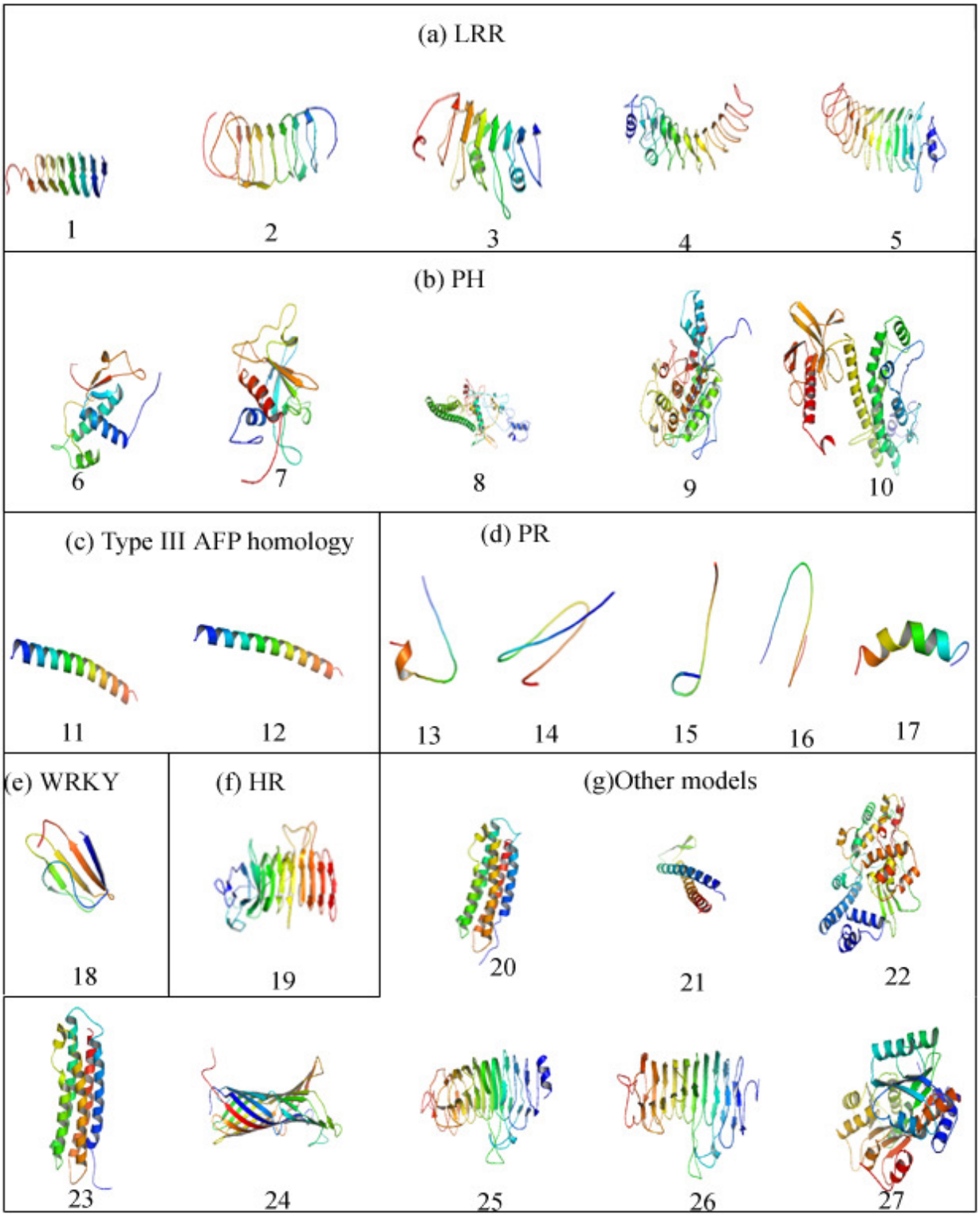



**Figure 2.** Snap shots of three-dimensional models of antifreeze proteins in over wintering plants. The 3D structures of antifreeze proteins were predicted through homology modeling and threading approaches. The structures were displayed by PyMOL visualization tool (http:// http://www.pymol.org/). Numerical order indicates: (a) Leucine rich repeat: (1) ACG75703, (2) ACM61985, (3) ACM61984, (4) Q56B90 and (5) Q56B89; (b) Pleckstrin homology: (6) Q6UAH5, (7) XP_002533988, (8) XP_002524729, (9) XP_002517076 and (10) XP_002516542; (c) Type III AFP homology: (11) AAB20142 and (12) AAB20141, (d) Pathogenesis related: (13) Q9S8C6, (14) Q9S8C7, (15) Q9S8C5, (16) Q9S8C8 and (17) Q9S899; (e) WRKY: (18) AAL26842, (f) Hemagglutinin related: (19) XP_002536221, and (g) Other models: (20) ABR01234, (21) ABR01231, (22) ABR01229, (23) ABR01227, (24) ABY64765, (25) ABY64764, (26) ABY64763 and (27) ABY64762.

**Table 5.** Model validation results of plant antifreeze proteins.

| Accession number | Template ID | Identity (%) | Approach used | Rampage - Ramachandran plot (%) | | | CE, RMSD (A°) | PMDB ID |
|---|---|---|---|---|---|---|---|---|
| | | | | Favored | Allowed | Outlier | | |
| AAB20142 | 1WFA | 100 | Homology modeling | 100 | 0 | 0 | 0.1 | PM0076159 |
| AAB20141 | 1WFA | 100 | Homology modeling | 100 | 0 | 0 | 0.1 | PM0076158 |
| Q9S8C6 | 1GHS | 100 | Homology modeling | 100 | 0 | 0 | 0.5 | PM0076160 |
| Q9S8C7 | 1Z3Q | 59 | Homology modeling | 96.3 | 0 | 3.7 | 0.5 | PM0076161 |
| Q9S8C5 | 2DKV | 85 | Homology modeling | 85.7 | 14.3 | 0 | 0.2 | PM0076162 |
| Q9S8C8 | 1DU5 | 61 | Homology modeling | 95.5 | 4.5 | 0 | 0.5 | PM0076163 |
| Q9S899 | 2GRV | 54 | Homology modeling | 100 | 0 | 0 | 0.2 | PM0076164 |
| Q6UAH5 | 1GD8 | LT 40 | Threading | 88.6 | 9.4 | 2 | 0.9 | PM0076165 |
| XP_002533988 | 1MAI | 42 | Homology modeling | 89.2 | 8.7 | 2.1 | 1.4 | PM0076166 |
| XP_002524729 | 1IEX, 1IH7, 1MAI, 1TXD, 2DA0, 2H94 | LT 40 | Threading | 81.0 | 14.5 | 4.5 | 1.2 | PM0076167 |
| XP_002517076 | 1PLS, 1N4C, 1D0V | LT 40 | Threading | 85.4 | 14.6 | 0 | 1.8 | PM0076168 |
| XP_002516542 | 1PLS, 1BTN, 1DRO, 1QQG, 1MR5 | LT 40 | Threading | 82.8 | 14.2 | 3 | 1.9 | PM0076169 |
| XP_002536221 | 1RWR | LT 40 | Threading | 89.7 | 7.3 | 3 | 0.8 | PM0076170 |
| ABR01234 | 1EQ1 | LT 40 | Threading | 90.3 | 6.8 | 2.9 | 0.7 | PM0076171 |
| ABR01231 | 2ZDI | 45 | Homology modeling | 96.1 | 2.9 | 1 | 1.1 | PM0076172 |
| ABR01229 | 1RQU, 1Y79 | LT 40 | Threading | 88 | 10.3 | 1.7 | 1.8 | * |
| ABR01227 | 1EQ1 | LT 40 | Threading | 91.5 | 5.7 | 2.8 | 0.7 | PM0076173 |
| ABY64765 | 1I78 | LT 40 | Threading | 88.5 | 8.7 | 2.8 | 1.3 | PM0076174 |
| ABY64764 | 1HG8 | LT 40 | Threading | 82.9 | 12.6 | 4.5 | 0.8 | PM0076175 |
| ABY64763 | 1HG8 | LT 40 | Threading | 83.9 | 10.5 | 5.6 | 0.8 | PM0076176 |
| ABY64762 | 1JX6 | 40 | Homology modeling | 94.9 | 4.3 | 0.8 | 1.7 | PM0076177 |
| ACG75703 | 1P9H | LT 40 | Threading | 88.5 | 10.6 | 0.9 | 0.3 | PM0076178 |
| ACM61985 | 1OGQ | 41 | Homology modeling | 86.1 | 11 | 2.9 | 0.4 | PM0076179 |
| ACM61984 | 2Z82 | 40 | Homology modeling | 91.6 | 7.8 | 0.6 | 1.8 | PM0076180 |
| Q56B90 | 1OGQ | LT 40 | Threading | 85.3 | 11.9 | 2.8 | 0.9 | PM0076181 |
| Q56B89 | 1OGQ | LT 40 | Threading | 85.3 | 11.9 | 2.8 | 0.9 | * |
| AAL26842 | 2AYD | 71 | Homology modeling | 100 | 0 | 0 | 0.2 | PM0076182 |

List of chosen templates with identity are mentioned. Twenty five refined models have been deposited into PMDB and accession codes are given. * indicates coordinates are also deposited and codes are pending for processing.

or bond length was detected in our refined models. Ramachandran Plot analysis revealed that, all the models were with good structural quality. The structural details and the refined three dimensional models of 27 AFPs have been deposited into the protein model data base (PMDB, http://mi.caspur.it/PMDB/) and their accession numbers are displayed in Table 5.

**LRR/IRI – AFPs**

LRR/IRI domain containing AFP models (*T. aestivum*: ACM61985, ACM61984, Q56B89, Q56B90 and *F. pratensis*: ACG75703) are conserved with already reported AFPs of *L. perenne and D. Antarctica*. As LRR, IRI domain is also involved in ice-recrystallization inhibitory activity. The 3D structure of AFP from *F. pratensis* is closely related to *L. perenne* - AFP, because the secondary structural identity of both the protein is 78%. Though LRR domain present in the AFPs, the predominant ice-binding region is predicted to be the IRI domain, and it contained two ice-binding surfaces, on either side of the β roll domain. The concave side of the predicted models is made up of continuous beta sheets and the convex side showed a variety of secondary structural elements such as helices and coil. Ascending and descending loops of predicted models made the connection between alpha helices and beta sheets as experimental LRR domain containing proteins. The structure classification of both experimental and most of the predicted LRR domain containing proteins is alpha + beta, which explains the reliability of the predicted models.

**PH - AFPs**

The PH domain contains 100 amino acid residues that occur in a broad range of proteins concerned in signal transduction mechanism or as constituents of the cytoskeleton. Several PH domain containing protein structures are solved with high resolution such as pleckstrin (N-terminal), β-spectrin, dynamin, phospholipase C-$\delta_1$ (PLC- $\delta_1$), Son of sevenless (Sos) and Bruton's tyrosine kinase (Btk) (Rebecchi and Scarlata, 1998). PH domains consist of two perpendicular anti-parallel beta sheets, followed by a C-terminal amphipathic helix. In our study, the four AFPs from *R. communis* are belonging to PH family (XP_002533988, XP_002524729, XP_002517076 and XP_002516542). However, the remaining protein (XP_002536221) was beta class, which is similar to hemagglutinin-related protein from *Granulibacter bethesdensis CGDNIH1*. The secondary structural class of our predicted PH domain containing AFP models is beta as experimental structure (Yoon et al., 1994). The PH domain containing proteins is not directly involved in ice-recrystallization inhibitory activity; however, it may be involved in intracellular signaling during the cold condition.

**WRKY - AFPs**

The WRKY proteins comprise major family of a transcription factor that is essential for plant disease resistance, abiotic stress, senescence and some developmental processes in plants. This kind of protein is also named as thermal hysteresis proteins (THP). Crystal structure of *Arabidopsis thaliana* WRKY1-C (C-terminal) domain is composed of five β- strands (β1: 294–300, β2: 312–318, β3: 327–332, β4: 340–345 and β 5, 352–358), which forms the antiparallel β- sheet. The zinc-binding site is located at one end of the β- sheet, between strands β4 and β5 (Yamasaki et al., 2005). WRKY domain containing AFP model from *S. dulcamara* is also consisted of five β- strands as experimental structure. The secondary structure of *At*WRKY1-C is beta. Our predicted WRKY AFP model is also coming under the same class. The function of *S. dulcamara* – AFP is intracellular, because it does not contain putative signal peptide sequence in N-terminal region. The main function for this protein is DNA-binding activity.

**Other AFPs**

The 3D structure of two AFPs (AAB20142 and AAB20141) from *N. tabacum* contain only alpha helices, which is exactly similar to the *Winter flounder* - antifreeze protein Isoform HPLC6 (1WFA). *S. cereale* – AFPs (Q9S8C6, Q9S8C7, Q9S8C5 and Q9S8C8) are generally termed as PR (Glucan endo-1, 3-beta-glucosidase, Thaumatin-like protein and Endochitinase) proteins, in which, their predominant secondary structural composition is predicted to be coil rather than helices and sheets. The composition of alpha helices (ABR01234: 74.16%, ABR01231: 68.27%, ABR01229: 70.19% and ABR01227: 71.35%) is high in *C. vulgaris* - AFPs, where as coil content (ABY64765: 59.5%, ABY64764: 61.84%, ABY64763: 61.80% and ABY64762: 61.76%) is high in *Chlamydomonas sp. CCMP681*- AFPs.

**Binding site analysis**

Get area was used to calculate the solvent accessible surface area/solvation energy, number of surface atoms and buried atoms of the refined models. The very high value of the solvation energy resulted in a better interaction with ice. Thematics web server was used to identify the potential binding site residues from the predicted models. This analysis can be used to study the surface features and functional regions of the refined models. The probable binding sites residues of all AFPs

**Table 6.** Solvent accessibility and predicted binding site residues of plant antifreeze proteins.

| S/No. | Accession number | Solvent accessibility | | | Probable binding residues of plant antifreeze proteins |
|---|---|---|---|---|---|
| | | Solvation energy/SASA | No. of surface atoms | No. of buried atoms | |
| 1 | AAB20142 | 2839.67 | 162 | 44 | ASP 4, THR 5, ALA 6, ASP 8, ALA 9 |
| 2 | AAB20141 | 2839.67 | 178 | 48 | ASP 2 , ASP 6 |
| 3 | Q9S8C6 | 2287.83 | 126 | 20 | TYR 5 , ARG 15 |
| 4 | Q9S8C7 | 2643.65 | 168 | 41 | LYS 8 |
| 5 | Q9S8C5 | 1681.51 | 90 | 16 | CYS 3 , CYS 12 |
| 6 | Q9S8C8 | 2640.72 | 152 | 23 | CYS 9 |
| 7 | Q9S899 | 2536.12 | 141 | 29 | VAL 1,GLU 9 |
| 8 | Q6UAH5 | 10477.17 | 786 | 450 | ARG 14 , LYS 18 , LYS 40 , CYS 41 , TYR 136 , HIS 138 , LYS 114 , ARG 112, ASP 129 |
| 9 | XP_002533988 | 11507.00 | 921 | 616 | GLU 46, ASP 53 , ASP 63 , GLU 95 , ASP 88, HIS 139 , ASP 179, HIS 185 |
| 10 | XP_002524729 | 31224.04 | 2532 | 1099 | ALA 420, TRP 421, PRO 422, THR 460, ASN 461, VAL 463, SER 464, LEU 479, GLN 482 |
| 11 | XP_002517076 | 20398.54 | 1710 | 1868 | GLU 356 , TYR 381 , LYS 404 , LYS 405 , ARG 432 |
| 12 | XP_002516542 | 22593.60 | 2002 | 1212 | LYS 263 , TYR 270, LYS 305 , CYS 346 , CYS 386 , LYS 376 , LYS 392 , TYR 372 |
| 13 | XP_002536221 | 13973.67 | 1309 | 852 | ASP 57 , ASP 77, HIS 93 , HIS 219, ASP 321 |
| 14 | ABR01234 | 9800.32 | 804 | 503 | ARG 12 , ARG 19 , GLU 128 , ASP 159, GLU 52 , ASP 53 , LYS 55 , ASP 56 , ASP 104 , GLU 174 , LYS 59, ASP 92 |
| 15 | ABR01231 | 7765.63 | 555 | 234 | ASP 50 , ASP 61, GLU 78, GLU 90 , HIS 93 , GLU 97 , ASP 96 |
| 16 | ABR01229 | 25524.65 | 2273 | 1305 | GLU 196 , GLU 267 , HIS 268 , LYS 233 , CYS 255 , GLU 257, LYS 303 , LYS 391 , GLU 445 , ASP 455 , ARG 384 , ARG 441, ARG 305, GLU 346, ARG 431 |
| 17 | ABR01227 | 9901.57 | 808 | 501 | GLU 52 , LYS 55 , ASP 56 , LYS 59 , ASP 96 , ASP 92, GLU 174 |
| 18 | ABY64765 | 19232.02 | 1669 | 900 | TYR 71 , CYS 147 , LYS 205 , TYR 277 , ARG 78 , ARG 292 , ARG 131, ASP 195 , ASP 241, LYS 27 , CYS 328 |
| 19 | ABY64764 | 15650.82 | 1496 | 1087 | ASP 71 , GLU 92 , GLU 117, ASP 310 , GLU 317 , CYS 339 , CYS 350 |

**Table 6.** Cont'd.

| | | | | | |
|---|---|---|---|---|---|
| 20 | ABY64763 | 15923.20 | 1505 | 1044 | ASP 91 , GLU 114, ASP 263 , GLU 302 , HIS 339 , CYS 303 , CYS 309 , TYR 312 , CYS 336 , CYS 347 |
| 21 | ABY64762 | 17157.43 | 1590 | 955 | GLU 54 , GLU 106, ASP 88, GLU 155 , ASP 286 , ASP 287 , ASP 238 , CYS 322 , LYS 334, LYS 296 , CYS 300 |
| 22. | ACG75703 | 6069.71 | 488 | 324 | HIS 54 , HIS 61 , ASP 65 , ASP 74 , ASP 73 |
| 23. | ACM61985 | 9148.36 | 859 | 423 | HIS 13, ARG 60 , ARG 61, ASP 91 , ASP 114, HIS 129, HIS 150 , HIS 171 , GLY 175, |
| 24. | ACM61984 | 11132.72 | 954 | 377 | CYS 45, LYS 93, TYR 110 , ARG 152 , ARG 155 , LYS 130 , ARG 176 |
| 25. | Q56B90 | 13709.62 | 1269 | 772 | ARG 19, ARG 68 , ARG 93 , HIS 117 , GLU 95 , CYS 119 , LYS 140 , TYR 120 , ASP 168 , ASP 122 , GLU 169 , HIS 147, ASP 235 |
| 26 | Q56B89 | 13709.62 | 1269 | 772 | ARG 19, ARG 68 , ARG 93 , HIS 117 , GLU 95 , CYS 119 , LYS 140 , TYR 120 , ASP 168 , ASP 122 , GLU 169 , HIS 147, ASP 235 |
| 27 | AAL26842 | 5119.16 | 389 | 205 | ASP 368, LYS 378, GLN 381, ASN 389, PRO 390, SER 392, TYR 394, VAL 407, GLU 408, VAL 417 |

were mentioned in Table 6. Based on the results of protein modeling and binding site analysis, these sites were chosen as the most favorable binding sites for further molecular docking studies of AFPs with ice. All the sequence and structural analysis helped into classifying the AFPs in over wintering plants (Figure 3), and it is essential information for further experimental studies.

## Conclusion

The present study involving analysis of various sequence and structural features of AFPs in over wintering plants has provided insight into the ice-recrystallization inhibition process. The dataset of 40 plant AFPs was retrieved from UniProt and GenPept databases and classify the AFPs into different categories. Primary structure analysis shows that most of AFPs are hydrophilic in nature. Computed pI value revealed that 21 AFPs posses the acidic property, whereas remaining to have basic property. The AFPs ACG75703, AAB20141, Q9M3W4 and AAB20142 cannot be used for UV spectrometry studies due to the absence of EC value. GRAVY index computes the possible crystallization propensity of all the AFPs, in which, twenty seven are more chances to get the crystal due to their high GRAVY score, whereas remaining were found to be optimum chances.

Molecular phylogenetic tree segregated the AFPs into various groups based on the sequence similarity and distances. The AFP from *F. pratensis* does not show any significant similarity with PGIPs. Moreover, transmembrane domains are observed in some of AFPs. *S. ceareale* and *P. suaveolens* AFPs (Q9S8C6 and Q6UAH5) do not contain glycosylation sites, whereas remaining AFPs were found to have several glycosylation sites. Presence of N-terminal signal peptide in AFPs suggests that these proteins would participate in the secretory pathway. In our study, 22 AFPs are targeted into extra cellular space and remaining is targeted into mitochondria, chloroplast and other locations. Our secondary structure results suggested that 7 AFPs are belonging to alpha, 15 is beta and 5 are alpha + beta

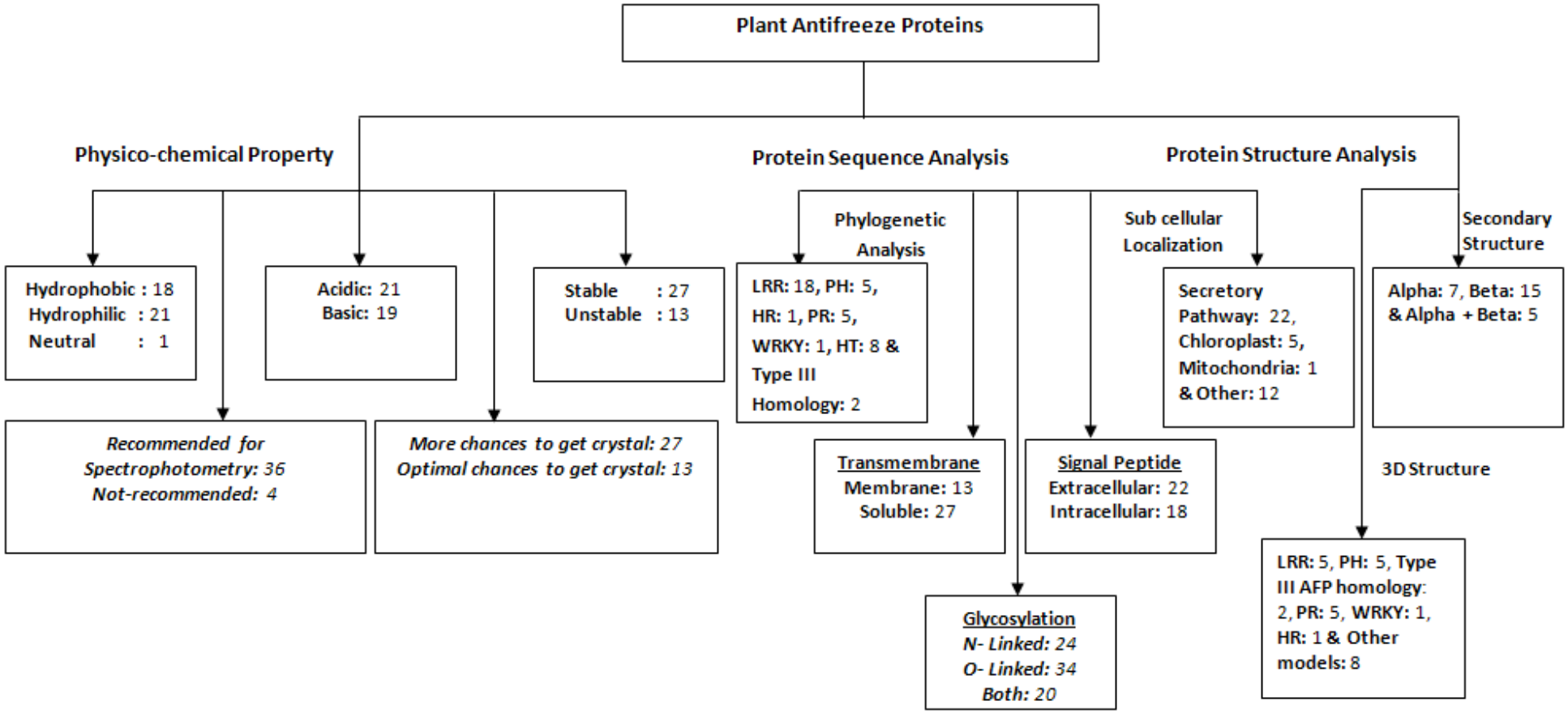



**Figure 3.** Classification of antifreeze proteins in overwintering plants**.** The plant antifreeze proteins were classified and distributed based on their sequence and structural features (LRR: Leucine rich repeat, PH: Pleckstrin homology, PR: Pathogenesis related, HR: Hemagglutinin related and HT: Hypothetical).

classes, respectively.

Ongoing efforts are directed towards the development of molecular docking and dynamics based CS towards plant AFPs to understand the mechanism of an ice-recrystallization inhibiton. Newly identified LRR, PR and HR AFPs can be validated by employing the molecular interactions studies with their interacting partners. The proposed classification scheme allowed us to identify the binding sites of AFP and design an engineered construct of potential AFP with superior ice-recrystallization inhibitory activity or in production of fusion protein, which will protect the plants from freezing conditions and psychrophilic pathogens. Multigene transformation could be necessary to transfer these characteristics to other plants. This type of research will eventually permit evaluation of the effectiveness of the AFPs for the enhancement of frost resistance of commercially important crops.

## ACKNOWLEDGEMENTS

J. Muthukumaran thanks Council for Scientific and Industrial Research (CSIR) for Senior Research Fellowship (SRF). P. Manivel thanks University Grant Commission (UGC), Government of India for providing financial assistance to carry out the research work. M. Kannan thanks UGC for Rajiv Gandhi National Fellowship to pursue his Ph.D degree. R. Krishna thanks Centre of Excellence in Bioinformatics, Pondicherry University funded by Department of Biotechnology and Department of Information technology, Government of India, New Delhi for providing the essential computational resources for carrying out the research work.

**Abbreviations: AFP,** Antifreeze proteins; **LRR,** Leucine rich repeat; **PH,** Pleckstrin homology; **PR,** Pathogenesis related; **PGIP,** Polygalatouronase inhibiting protein; **SMART,** Simple modular architecture research tool; **MSA,** Multiple sequence alignment; **LRRNT,** Leucine rich repeat N-terminal; **RI,** Recrystallization inhibition; **TH,** thermal hysteresis; **GRAVY,** grand average of hydropathicity; **PCoA,** principal coordinate analysis; **PDB,** protein data bank; **SPC,** single point charge; **CE,** combinatorial extension; **THEMATICS**, Theoretical macroscopic titration curves; **EC,** extinction coefficient; **pI,** isoelectric point; **NJ,** neighbor-joining; **RMSD,** root mean square deviation; **PMDB,** protein model data base.

### REFERENCES

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402.

Bayer-Giraldi M, Uhlig C, John U, Mock T and Valentin K (2010). Antifreeze proteins in polar sea ice diatoms: diversity and gene expression in the genus Fragilariopsis. Environ. Microbiol. 12: 1041-52.

Brudno M, Steinkamp R, Morgenstern B (2004). The CHAOS/DIALIGN WWW server for multiple alignment of genomic sequences. Nucleic Acids Res. 32: W41-44.

Choi JH, Jung HY, Kim HS, Cho HG (2000). PhyloDraw: a phylogenetic tree drawing system. Bioinformatics 16: 1056-1058.

Davies PL, Baardsnes J, Kuiper MJ, Walker VK (2002). Structure and function of antifreeze proteins. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 357: 927-935.

Deng G, Andrews DW, Laursen RA (1997). Amino acid sequence of a new type of antifreeze protein, from the longhorn sculpin Myoxocephalus octodecimspinosis. FEBS Lett. 402: 17-20.

Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. 2: 953-971.

Eswar N, Eramian D, Webb B, Shen MY, Sali A (2008). Protein structure modeling with MODELLER. Methods Mol. Biol. 426: 145-159.

Fletcher GL, Hew CL, Davies PL (2001). Antifreeze proteins of teleost fishes. Annu. Rev. Physiol. 63: 359-390.

Fraczkiewics RB, Braun W (1998). Exact and efficient analytical calculation of the accessible surface areas and their Gradients for Macromolecules. J. Comp. Chem 19: 319-333.

Raghava GPS (1999). Proclass: A computer program for predicting the protein structural classes. J. Biosciences 24: 176.

Griffith M, Yaish MW (2004). Antifreeze proteins in overwintering plants: a tale of two activities. Trends Plant Sci. 9: 399-405.

Hess B (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J. Chem. Theory. Comput 4: 435-447.

Hirokawa T, Boon-Chieng S, Mitaku S (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics 14: 378-379.

Hon WC, Griffith M, Mlynarz A, Kwok YC, Yang DS (1995). Antifreeze proteins in winter rye are similar to pathogenesis-related proteins. Plant Physiol. 109: 879-889.

Huang T, Duman JG (2002). Cloning and characterization of a thermal hysteresis (antifreeze) protein with DNA-binding activity from winter bittersweet nightshade, Solanum dulcamara. Plant Mol. Biol. 48: 339-350.

John UP, Polotnianka RM, Sivakumaran KA, Chew O, Mackin L, Kuiper MJ, Talbot JP, Nugent GD, Mautord J, Schrauf GE, Spangenberg GC (2009). Ice recrystallization inhibition proteins (IRIPs) and freeze tolerance in the cryophilic Antarctic hair grass Deschampsia antarctica E. Desv. Plant Cell Environ. 32: 336-348.

Kelley LA, Sternberg MJ (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. 4: 363-371.

Kuiper MJ, Davies PL, Walker VK (2001). A theoretical model of a plant antifreeze protein from Lolium perenne. Biophys. J. 81: 3560-3565.

Letunic I, Doerks T, Bork P (2009). SMART 6: recent updates and new developments. Nucleic Acids Res. 37: D229-232.

Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins 50: 437-450.

Meyer K, Keil M, Naldrett MJ (1999). A leucine-rich repeat protein of carrot that exhibits antifreeze activity. FEBS Lett. 447: 171-178.

Ng NF, Hew CL (1992). Structure of an antifreeze polypeptide from the sea raven. Disulfide bonds and similarity to lectin-binding proteins. J. Biol. Chem. 267: 16069-16075.

Ondrechen MJ, Clifton JG, Ringe D (2001). THEMATICS: a simple computational predictor of enzyme function from structure. Proc. Natl. Acad. Sci. USA 98: 12473-12478.

Raymond JA, Janech MG, Fritsen CH (2009). Novel ice-binding proteins from a psychorophilic antarctic alga (Chlamydomonadaceae, Chlorophyceae). J. Phycol. 45: 130-136.

Rebecchi MJ, Scarlata S (1998). Pleckstrin homology domains: a common fold with diverse functions. Ann. Rev. Biophys. Biomol. Struct. 27: 503-528.

Retief JD (2008). Phylogenetic Analysis using PHYLIP. In: Krawetz, S.M.a.S.A. (Ed.), Bioinformatics Methods and Protocols, Humana Press, pp. 243-258.

Scotter AJ, Marshall CB, Graham LA, Gilbert JA, Garnham CP, Davies PL (2006). The basis for hyperactivity of antifreeze proteins.

Cryobiology 53: 229-239.
Sen TZ, Jernigan RL, Garnier J, Kloczkowski A (2005). GOR V server for protein secondary structure prediction. Bioinformatics 21: 2787-2788.
Shindyalov IN, Bourne PE (1998). Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng. 11: 739-747.
Sicheri F, Yang DS (1995). Ice-binding structure and mechanism of an antifreeze protein from winter flounder. Nature 375: 427-431.
Sidebottom C, Buckley S, Pudney P, Twigg S, Jarman C, Holt C, Telford J, McArthur A, Worrall D, Hubbard R, Lillford P (2000). Heat-stable antifreeze protein from grass. Nature 406: 256.
Sonnichsen FD, Sykes BD, Chao H, Davies PL (1993). The nonhelical structure of antifreeze protein type III. Science 259: 1154-1157.
Thomashow MF (1998). Role of cold-responsive genes in plant freezing tolerance. Plant Physiol. 118: 1-8.
Tyagi N, Anamika K, Srinivasan N (2010). A framework for classification of prokaryotic protein kinases. PLoS One. 26: e10608.
Venketesh S, Dayananda C (2008). Properties, potentials, and prospects of antifreeze proteins. Crit. Rev. Biotechnol. 28: 57-82.
Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF(1999). Protein identification and analysis tools in the ExPASy server. Methods Mol. Biol. 112: 531-552.
Worrall D, Elias L, Ashford D, Smallwood M, Sidebottom C, Lillford P, Telford J, Holt C, Bowles D (1998). A carrot leucine-rich-repeat protein that inhibits ice recrystallization. Science 282: 115-117.
Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N, Suzek B (2006). The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Res. 34: D187-191.
Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Tomo Y, Hayami N, Terada T, Shirouzu M, Tanaka A, Seki M, Shinozaki K, Yokoyama S (2005). Solution structure of an Arabidopsis WRKY DNA binding domain. Plant Cell 17: 944-956.
Yoon HS, Hajduk PJ, Petros AM, Olejniczak ET, Meadows RP, Fesik SW (1994). Solution structure of a pleckstrin-homology domain. Nature 369: 672-675.
Zhang DQ, Liu B, Feng DR, He YM, Wang SQ, Wang HB, Wang JF (2004). Significance of conservative asparagine residues in the thermal hysteresis activity of carrot antifreeze protein. Biochem. J. 377: 589-595.