nature biotechnology

Ordered surface carbons distinguish antifreeze proteins and their ice-binding regions

Andrew C Doxey[1], Mahmoud W Yaish[1,2], Marilyn Griffith[1] & Brendan J McConkey[1]

Antifreeze proteins (AFPs) are found in cold-adapted organisms and have the unusual ability to bind to and inhibit the growth of ice crystals. However, the underlying molecular basis of their ice-binding activity is unclear because of the difficulty of studying the AFP-ice interaction directly and the lack of a common motif, domain or fold among different AFPs. We have formulated a generic ice-binding model and incorporated it into a physicochemical pattern-recognition algorithm. It successfully recognizes ice-binding surfaces for a diverse range of AFPs, and clearly discriminates AFPs from other structures in the Protein Data Bank[1]. The algorithm was used to identify a novel AFP from winter rye, and the antifreeze activity of this protein was subsequently confirmed. The presence of a common and distinct physicochemical pattern provides a structural basis for unifying AFPs from fish, insects and plants.

Understanding how AFPs bind to ice has been a fundamental question in AFP research since their discovery in fish and insects over 30 years ago[2] and is crucial in the development of potential biotechnological applications[3] such as cryogenic storage, cryosurgery, food preservation, freeze-resistant crops and rational AFP design. However, since AFPs cannot be crystallized with ice[4], there is no direct approach for identifying AFP-ice interfaces. Thus, structure-function studies rely mainly on site-directed mutagenesis and computational methods, which have led to several hypotheses regarding AFP-ice interactions. To determine what distinguishes AFPs from the ~99.9% of proteins that have no observable affinity for ice[4], a generic model could be used to compare a full set of AFP structures with other proteins in the Protein Data Bank (PDB). One interaction model proposes that AFPs bind ice via relatively flat, hydrophobic surfaces that are complementary to ice surfaces[4]. Testing such a model, however, is not feasible using current algorithmic methods (e.g., molecular docking) as they are computationally intensive for just a single AFP-ice system and are highly dependent on the choice of energy function and the definition of the ice surface[5]. To circumvent this issue, we have developed a rapid algorithm to score proteins directly based on a set of physicochemical surface features, allowing for comparative evaluation of multiple AFPs and thousands of PDB structures.

The basic tenet of our algorithm is that spatially regular surface atoms should have an increased probability of docking to an ordered substrate such as ice and, therefore, should be most concentrated on the ice-binding surface. Spatial regularity is scored explicitly by computing vectors between solvent-accessible atoms and determining which atoms are associated with a local, repetitive vector pattern (**Fig. 1**). This approach will recognize highly ordered, relatively planar surfaces stemming from a repetitive geometric pattern, thus allowing for detection of ice-binding surfaces regardless of their ice-plane specificity and binding orientation. Extending the 'hydrophobic surface' model[4,6], the algorithm assesses the spatial regularity of solvent-accessible, nonpolar carbons instead of polar atoms. This is consistent with an ice-binding mechanism in which carbons (e.g., methyl groups) penetrate regular spacings (e.g., ice cages, grooves, hexagonal rings) on the ice surface, forming favorable van der Waals contacts[5,7]. Carbons from aromatic and charged residues have been omitted, as aromatic groups are too bulky to be accommodated within interstitial spaces of ice and have been shown by mutational studies[8] to reduce antifreeze activity, and interactions of charged residues energetically favor liquid water over ice[9]. An additional desirable feature of an ice-binding surface is that it is sufficiently hydrophilic to be exposed to solvent in its native state. Therefore, only surface carbons in close proximity to polar atoms are selected. According to this model, polar atoms are required primarily for solubility, but a supplementary role

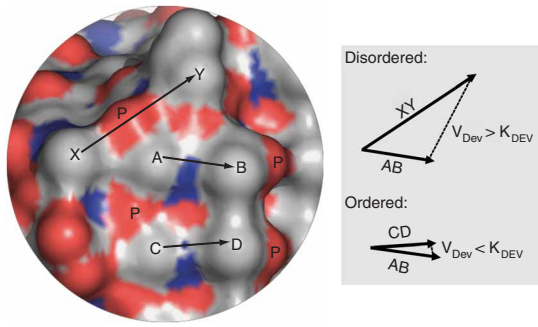

Figure 1 Identification of ordered surface carbons using vector comparison algorithm. Atoms {A,B,C,D,X,Y} are sufficiently solvent-exposed carbons (SAS > K_{SAS}). Atoms {A,B,C,D} are spatially regular since |AB - CD| < K_{Dev} and {A,B} is near {C,D} (within distance K_{HIGH}). These atoms are in also close proximity (within distance K_{polar}) to polar atoms (some of these have been labeled P). {A,B,C,D} are ordered surface carbons (OSCs).

[1]Department of Biology, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada. [2]Current address: Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario, N1G 2W1 Canada. Correspondence should be addressed to B.J.M. (mcconkey@uwaterloo.ca).

Received 4 August 2005; accepted 5 May 2006; published online 2 July 2006; doi:10.1038/nbt1224

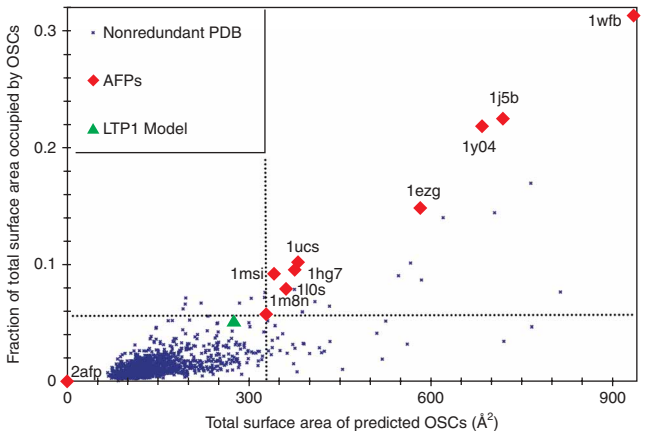

Figure 2 Scoring of AFPs versus nonredundant PDB structures. Each point represents a single structure. The structures are positioned along the *y*-axis by one scoring scheme (size of predicted ice-binding surface as fraction of surface area) and along the *x*-axis by another scoring scheme (size of predicted ice-binding surface, independent of total surface area). The plot has been divided in four quadrants at the data point representing structure 1m8n. The two left quadrants contain one AFP structure (type II NMR model) and 3,165 others. The bottom-right quadrant contains no AFPs and 29 others, and the top right quadrant contains nine AFPs (seven X-ray structures and two NMR models) and 12 others.

in hydrogen-bonding interactions with ice cannot be ruled out. The final output of the algorithm for a putative AFP is an identified set of spatially regular, hydrophobic carbons, termed ordered surface carbons (OSCs).

We applied the algorithm to a representative set of AFP structures from the PDB, including three each of type I AFPs, type III AFPs and β-helical insect AFPs, as well as the one available NMR model of type II AFP (see ref. 4 for a summary of known antifreeze types). Predicted OSCs (**Supplementary Table 1** and **Supplementary Fig. 1** online) for nearly all AFPs were in excellent agreement with ice-binding residues identified by previous studies[6,10,11]. Because of the structural regularity and amphipathic nature of alpha-helical type I AFPs, OSCs were found predominantly on their conserved, alanine-rich, 'hydrophobic faces'[6]. Moreover, OSCs detected on insect AFPs corresponded to their highly ordered, threonine faces, and OSCs recognized on type III AFP were in almost perfect correspondence with established ice-binding residues[10]. Previous type III AFP ice-binding models involve many of these residues in H-bonding interactions with ice[12–14]. We have shown, however, that these sites can be detected by focusing on OSCs and that consideration of the orientation of polar atoms is not required. This is further reinforced by the relatively small reduction in performance when polar atom information is omitted (**Supplementary Table 2** online).

In many cases, the identified ice-binding surfaces represent relatively flat regions on the protein surface, an important feature of ice binding that was used by the 'flatness function' algorithm[7]. In addition to planar residues, the vector-based algorithm presented here also identified out-of-plane residues flanking the planar region of the ice-binding surface, which have been implicated in type III AFP's mechanism of action[10,13]. Such peripheral residues may allow type III AFP to simultaneously bind multiple ice-surfaces[4] or may alter hydration structure to further impede crystal growth[13].

To determine whether AFPs could be distinguished from other structures in the PDB based on algorithmic prediction of OSCs, 3,206 nonredundant structures including ten wild-type AFPs were scored (**Fig. 2**). When AFPs were scored by fraction of surface area occupied by predicted OSCs (FSA), nine out of ten ranked above 99.3% of other PDB structures. Scored by total surface area of predicted OSCs (TSA) alone, these AFPs ranked above the 98.6% of other PDB structures. Type III mutant structures also ranked very highly relative to other structures in the PDB, with an average rank of 99.3% by FSA and 93.8% by TSA. In addition, when the algorithm stringency was relaxed to account for lower resolution models in the type III mutant data set, a correlation ($r = 0.66$) emerged between algorithm score and thermal hysteresis (**Supplementary Fig. 2** online). This is consistent with a previous neural-network study, which found nonpolar solvent–accessible surface area to be most closely correlated with thermal hysteresis[13].

Discriminative ability and ice-binding surface prediction accuracy were generally lower for NMR models compared with X-ray structures. However, whereas the representative type II NMR model received a score of zero, one of the five models within PDB ID 2AFP received high scores of 96.2% (FSA) and 94.5% (TSA). Variable results for NMR models are not surprising as the algorithm relies on interatomic distances and parameters on the order of 1–2 Å, an atomic resolution not typically achieved using NMR. An inability to predict protein interaction sites using NMR data has been reported previously[15], and was also attributed to insufficient resolution of surface side chains. In these cases, relaxing the parameters (that is, lowering the stringency of vector matching) can improve identification of true OSCs but also increases the rate of false positives. Other highly scored structures in addition to AFPs included viral coat proteins, membrane proteins and subunits within protein complexes. In these cases, putative ice-binding surfaces detected by the algorithm are likely to be internal interfaces and would be unable to interact with ice.

As there were no obvious novel candidate AFPs within the top ranked structures, we applied the algorithm to homology models based on a set of sequences previously isolated from the freezing-tolerant plants winter rye (*Secale cereale*) and *Thellungiella salsuginea*. The isolated sequences included lipid transfer protein (LTP) homologs, some of which are known to be expressed in response to cold and have cryoprotective functions[16] but have not been shown to have antifreeze activity. Where possible, structural models were generated using X-ray structure templates. From twenty constructed models, a lipid transfer protein homolog (LTP1) yielded a significant ice-binding surface score and was selected as a positive test. The algorithm recognized a set of six highly planar OSCs (four methyl groups from alanine and threonine residues and two β-carbons from cysteine and asparagine residues) forming the putative ice-binding surface. Scored among structures in the nonredundant set, LTP1 ranked higher than 99.2% of other PDB structures using FSA and 97.5% using TSA. We considered LTP1's score to indicate a high probability of ice-binding activity, particularly given that a homology model was used. Another modeled lipid transfer protein homolog (LTP2) received a score of zero despite having 70% sequence identity with LTP1 (**Fig. 3a**). The lower score relative to LTP1 resulted from a small number of amino acid substitutions on the analogous surface (**Fig. 3b**), which presented LTP2 as an ideal negative test. In place of several predicted ice-binding residues in LTP1 (Thr3, Ala44, and Thr48) are Ser3, Asp44 and Ala48, which disrupt regularity and introduce a charged side-chain to the putative ice-binding surface.

The predicted ice-binding activity was tested experimentally using the ice-crystal growth morphology assay[17–19] (**Fig. 3c**). LTP1 bound to ice, producing hexagonal shapes (characteristic of AFP-ice interaction) that grew further to produce unusual six-pointed-star–shaped crystals as the temperature was decreased. Similar morphologies have been observed in isolates from antifreeze-active fungi[20]. In contrast, LTP2

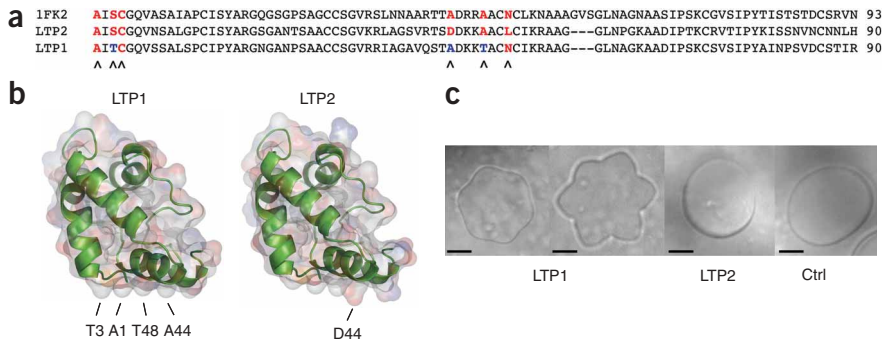

Figure 3 Prediction of ice-binding activity from homology models. (**a**) Sequence alignment of LTP1, LTP2 and template 1FK2. Marked residues correspond to predicted ice-binding sites in LTP1. Key substitutions are colored in blue. (**b**) Structural models of LTP1 and LTP2, visualized using PyMOL (http://www.pymol.org). Both structures have identical folds but differ in the degree of spatial regularity of the predicted ice-binding surface region. Several OSCs in LTP1 are labeled as the corresponding residue and the key substitution A44D is labeled on LTP2. The putative ice-binding surface of LTP1 is perpendicular to the plane of this figure. (**c**) Antifreeze activity of recombinant LTPs by ice-crystal morphology assay. Ice-crystals in LTP1 solution (labeled LTP1) initially grew as hexagons (left), and developed into six-pointed stars (right) as the temperature was cooled and warmed in very slow cycles. No antifreeze activity was observed for the control sample or LTP2 as evidenced by round, flat ice-crystals (labeled LTP2 and Ctrl). Incubation with proteinase K eliminated LTP1 antifreeze activity. Scale bars represent 10 μm.

did not exhibit ice-binding activity as evidenced by a round, flat ice-crystal morphology, which was indistinguishable from control samples. The ability of the algorithm to predict ice-binding activity for a new AFP fold provides additional experimental validation for our model. Interestingly, LTP1 is cysteine-rich and contains four disulfide bridges. High disulfide content is a characteristic of several AFP folds including β-helical insect and type II AFPs and has been proposed to provide a rigid conformation for enhanced stability at low temperatures[21]. Because the LTP sequences were identified in cold-acclimated winter rye, ice-binding is quite likely to be related to the biological function of LTP1 and may be an important feature of some other LTPs. For example, the LTP homolog cryoprotectin is known to be cold induced and is thought to prevent rupture of chloroplast thylakoids at freezing temperature[16,22]. Therefore, the ability to bind ice and affect crystal growth may not be specific to LTP1, but may also offer insight into the function of additional LTPs within this large and incompletely characterized protein family.

As in the case of AFPs, substrate-binding sites can be diverse in sequence and even fold. For these situations, a physicochemical surface pattern may better represent potential binding activity. Using this approach, we have computationally determined a set of features unique to AFPs from fish, insects and plants, which may provide insight into the nature of the AFP-ice interaction. Because the algorithm is not limited to particular vector orientations and surface geometries, it applies to a wide range of AFPs with different ice-binding surfaces, binding orientations, ice-plane specificities and resulting ice-crystal growth morphologies. Future adaptations of the algorithm that include more detailed analysis of vector-geometry may prove useful in detecting additional details of ice-binding such as these. Outside the realm of AFP research, the general approach of vector-based surface analysis may be extended to develop and test hypotheses regarding structure-function relationships of other protein classes.

METHODS

Algorithm. The algorithm (see **Supplementary Fig. 3** online) was developed in the Perl programming language, and uses a standard PDB file as input. Per-atom solvent-accessible surface area (SAS) is calculated using a rapid Voronoi procedure[23]. Sufficiently solvent-accessible carbons are isolated using a cutoff value (K_{SAS} = 15.1 Å^2). Carbons from uncharged, nonaromatic amino acids within distance K_{POLAR} (4 Å) of a solvent-accessible polar atom (oxygen or nitrogen) are selected. Two carbon pairs {A,B} and {C,D} meeting the above criteria are scored for regularity if the minimum distance between atoms in separate pairs is within K_{LOW} to K_{HIGH}. These atoms are deemed ordered if they produce equivalent vectors ($|\mathbf{AB}\text{-}\mathbf{CD}| < K_{DEV}$). OSCs from equivalent vectors are clustered together if their minimum interatomic distance is within distance K_{HIGH}. The largest cluster of OSCs is defined as the putative ice-binding surface. A quantitative measure of AFP character is estimated as the surface area occupied by these carbons. Both fractional and total surface areas are calculated. As carbons should also be adequately separated to prevent steric interference with ice-binding, an upper bound (K_{HIGH} = 7.66 Å) and lower bound (K_{LOW} = 4.50 Å) limits the length of carbon-to-carbon vectors. K_{DEV} (1.09 Å), the maximum deviation allowed for two vectors to be considered equivalent, controls the stringency of the algorithm, and is optimized based on the resolution of target structures. Parameter values were initially set based on theoretical considerations (that is, K_{HIGH} and K_{LOW} correspond closely to the unit cell dimensions of hexagonal ice (7.345 Å and 4.516 Å)[24]), and were locally adjusted to increase stringency while maintaining ice-binding surface prediction accuracy. Three previously characterized AFP structures (PDB IDs 1WFA, 1EZG and 1MSI) were used to validate parameter values.

Three-dimensional structure set. A set of 3,196 nonredundant structures was generated using the PISCES server[25]. Protein chains between 40 and 10,000 residues with less than 25% pairwise sequence identity, R-factors <0.3 and with resolution <3Å were selected. NMR entries, C-alpha-only structures and error-containing PDB files were excluded to obtain a set of high quality X-ray structures. A representative set of AFP structures from the PDB was obtained by selecting the highest resolution structures (or most representative NMR model) from separate crystallization or NMR studies, and then selecting the top three (if available) structures for separate AFP classes (type I, type II, type III and beta-helical insect AFPs). The most representative model for NMR studies was determined using the NMRCLUST procedure from the OLDERADO (On-Line Database Ensemble Representatives and Domains)[26]. AFP structures not already in the nonredundant set were added manually.

AFP prediction and homology modeling. A set of protein sequences was obtained in previous studies of freezing-tolerant plants winter rye and *T. salsuginea* (unpublished data). All sequences were identified from cold-acclimated plant tissue but the corresponding proteins had not been purified or tested for ice-binding activity. Of these, twenty novel sequences were homologous to proteins of known structure and permitted comparative modeling. Three-dimensional models were constructed using a fully automated procedure available through the SWISS-MODEL server[27]. No further manipulations were made to the models. Entries with the lowest *E*-value in a PDB-BLAST search were selected as structural templates for homology modeling. Nonspecific lipid-transfer protein from maize (PDB ID 1FK2) was used as a template for both LTP1 and LTP2 models. Swiss PDB Viewer version 3.7 was used for structure visualization[28], and WHATCHECK[29] was used for structure verification and revealed no major errors.

Cloning of lipid transfer proteins from winter rye. LTP1 and LTP2 cDNA clones were inserted in the pGEX-KG expression vector and fused with GST tags. Positive colonies were isolated, and the recombinant fused protein in the

lysate was purified using 1 ml GSTrap FF and HiTrap Benzamidine FF columns (Amersham Biosciences) according to established protocols.

Antifreeze assays. Antifreeze activity was assayed qualitatively by observing ice-crystal growth morphology in solution using a Clifton nanoliter osmometer mounted on a phase-contrast photomicroscope (Olympus BHT). Solutions were flash frozen and melted until a single ice-crystal remained, and cooled to observe changes in morphology as the ice crystal grew. Protein concentrations were 33 μg/ml. Multifaceted (e.g., hexagonally shaped) ice crystals are produced in the presence of AFPs, whereas round, flat ice-crystals indicate lack of ice-binding activity[17–19]. Protease treatment was used to verify that crystal morphology was due to the presence of protein, as described previously[30].

Accession codes. The mRNA sequences corresponding to LTP1 and LTP2 have been deposited in GenBank with accession numbers DQ641934 and DQ641935 respectively.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We would like to dedicate this publication to the memory of Marilyn Griffith. This work was supported by grants from the National Science and Engineering Research Council (NSERC) to B.J.M. and M.G., and we acknowledge the Killam Research Fellowship Program on behalf of M.G. We thank Barbara Moffatt for providing *T. salsuginea* sequences and Bernard Duncker for comments on the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturebiotechnology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Berman, H.M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
2. Duman, J. & DeVries, A. Freezing resistance in winter flounder *Pseudopleuronectes americanus*. *Nature* **247**, 237–238 (1974).
3. Knight, C.A. Adding to the antifreeze agenda. *Nature* **406**, 249–251 (2000).
4. Jia, Z. & Davies, P.L. Antifreeze proteins: an unusual receptor-ligand interaction. *Trends Biochem. Sci.* **27**, 101–106 (2002).
5. Jorov, A., Zhorov, B.S. & Yang, D.S. Theoretical study of interaction of winter flounder antifreeze protein with ice. *Protein Sci.* **13**, 1524–1537 (2004).
6. Harding, M.M., Ward, L.G. & Haymet, A.D. Type I 'antifreeze' proteins. Structure-activity studies and mechanisms of ice growth inhibition. *Eur. J. Biochem.* **264**, 653–665 (1999).
7. Yang, D.S. *et al.* Identification of the ice-binding surface on a type III antifreeze protein with a "flatness function" algorithm. *Biophys. J.* **74**, 2142–2151 (1998).
8. DeLuca, C.I., Davies, P.L., Ye, Q. & Jia, Z. The effects of steric mutations on the structure of type III antifreeze protein and its interaction with ice. *J. Mol. Biol.* **275**, 515–525 (1998).
9. Gallagher, K.R. & Sharp, K.A. Analysis of thermal hysteresis protein hydration using the random network model. *Biophys. Chem.* **105**, 195–209 (2003).
10. Antson, A.A. *et al.* Understanding the mechanism of ice binding by type III antifreeze proteins. *J. Mol. Biol.* **305**, 875–889 (2001).
11. Graether, S.P. & Sykes, B.D. Cold survival in freeze-intolerant insects: the structure and function of beta-helical antifreeze proteins. *Eur. J. Biochem.* **271**, 3285–3296 (2004).
12. Chao, H., Sonnichsen, F.D., DeLuca, C.I., Sykes, B.D. & Davies, P.L. Structure-function relationship in the globular type III antifreeze protein: identification of a cluster of surface residues required for binding to ice. *Protein Sci.* **3**, 1760–1769 (1994).
13. Graether, S.P. *et al.* Quantitative and qualitative analysis of type III antifreeze protein structure and function. *J. Biol. Chem.* **274**, 11842–11847 (1999).
14. Jia, Z., DeLuca, C.I., Chao, H. & Davies, P.L. Structural basis for the binding of a globular antifreeze protein to ice. *Nature* **384**, 285–288 (1996).
15. Fernandez-Recio, J., Totrov, M., Skorodumov, C. & Abagyan, R. Optimal docking area: a new method for predicting protein-protein interaction sites. *Proteins* **58**, 134–143 (2005).
16. Hincha, D.K. *et al.* Cabbage cryoprotectin is a member of the nonspecific plant lipid transfer protein gene family. *Plant Physiol.* **125**, 835–846 (2001).
17. DeVries, A.L. Antifreeze glycopeptides and peptides: interactions with ice and water. *Methods Enzymol.* **127**, 293–303 (1986).
18. Griffith, M. & Yaish, M.W. Antifreeze proteins in overwintering plants: a tale of two activities. *Trends Plant Sci.* **9**, 399–405 (2004).
19. Hon, W.C., Griffith, M., Chong, P. & Yang, D. Extraction and isolation of antifreeze proteins from winter rye (*Secale cereale* L.) leaves. *Plant Physiol.* **104**, 971–980 (1994).
20. Snider, C., Hsiang, T., Zhao, G. & Griffith, M. Role of ice nucleation and antifreeze activities in pathogenesis and growth of snow molds. *Phytopathology* **90**, 354–361 (2000).
21. Gauthier, S.Y., Kay, C.M., Sykes, B.D., Walker, V.K. & Davies, P.L. Disulfide bond mapping and structural characterization of spruce budworm antifreeze protein. *Eur. J. Biochem.* **258**, 445–453 (1998).
22. Sieg, F., Schroder, W., Schmitt, J.M. & Hincha, D.K. Purification and characterization of a cryoprotective protein (cryoprotectin) from the leaves of cold-acclimated cabbage. *Plant Physiol.* **111**, 215–221 (1996).
23. McConkey, B.J., Sobolev, V. & Edelman, M. Quantification of protein surfaces, volumes and atom-atom contacts using a constrained Voronoi procedure. *Bioinformatics* **18**, 1365–1373 (2002).
24. Madura, J.D., Baran, K. & Wierzbicki, A. Molecular recognition and binding of thermal hysteresis proteins to ice. *J. Mol. Recognit.* **13**, 101–113 (2000).
25. Wang, G. & Dunbrack, R.L., Jr. PISCES: a protein sequence culling server. *Bioinformatics* **19**, 1589–1591 (2003).
26. Kelley, L.A. & Sutcliffe, M.J. OLDERADO: on-line database of ensemble representatives and domains. *Protein Sci.* **6**, 2628–2630 (1997).
27. Schwede, T., Kopp, J., Guex, N. & Peitsch, M.C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* **31**, 3381–3385 (2003).
28. Guex, N. & Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997).
29. Hooft, R.W., Vriend, G., Sander, C. & Abola, E.E. Errors in protein structures. *Nature* **381**, 272 (1996).
30. Bravo, L.A. & Griffith, M. Characterization of antifreeze activity in Antarctic plants. *J. Exp. Bot.* **56**, 1189–1196 (2005).