

Quantitative and Qualitative Analysis of Type III Antifreeze Protein Structure and Function*

(Received for publication, December 23, 1998, and in revised form, January 26, 1999)

Steffen P. Graether, Carl I. DeLuca‡, Jason Baardsnes, Gregory A. Hill, Peter L. Davies, and Zongchao Jia§

From the Department of Biochemistry, Queen's University, Kingston, Ontario, K7L 3N6 Canada

Some cold water marine fishes avoid cellular damage because of freezing by expressing antifreeze proteins (AFPs) that bind to ice and inhibit its growth; one such protein is the globular type III AFP from eel pout. Despite several studies, the mechanism of ice binding remains unclear because of the difficulty in modeling the AFP-ice interaction. To further explore the mechanism, we have determined the x-ray crystallographic structure of 10 type III AFP mutants and combined that information with 7 previously determined structures to mainly analyze specific AFP-ice interactions such as hydrogen bonds. Quantitative assessment of binding was performed using a neural network with properties of the structure as input and predicted antifreeze activity as output. Using the cross-validation method, a correlation coefficient of 0.60 was obtained between measured and predicted activity, indicating successful learning and good predictive power. A large loss in the predictive power of the neural network occurred after properties related to the hydrophobic surface were left out, suggesting that van der Waal's interactions make a significant contribution to ice binding. By combining the analysis of the neural network with antifreeze activity and x-ray crystallographic structures of the mutants, we extend the existing ice-binding model to a two-step process: 1) probing of the surface for the correct ice-binding plane by hydrogen-bonding side chains and 2) attractive van der Waal's interactions between the other residues of the ice-binding surface and the ice, which increases the strength of the protein-ice interaction.

Many poikilothermic organisms have developed antifreeze proteins (AFPs)¹ to resist freezing. Five classes of structurally diverse antifreeze proteins have been found in fish (for a review, see Ref. 1). These proteins act by adsorbing to the surface of ice and increasing the curvature of ice fronts between the bound AFPs (2). The freezing point at the surface is depressed,

* This work was supported by the Medical Research Council of Canada. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors Gln9Thr/Gln44Thr, 3ame; Asn14Gln, 2ame; Asn14Ser/Ala16His, 8ame; Asn14Ser/Gln44Thr, 8msi; Thr15Ala, 7ame; Thr15Ser, 2spg; Thr15Val, 1msj; Thr18Ala, 4ame; Thr18Asn, 9msi; Thr18Ser, 1jab; Val20Ala, 1b7j; Met21Ala, 6ame; Glu35Lys, 1ekl; Ser42Gly, 9ame; Asn46Ser, 2msj; Arg47His, 1b7k; Lys61Arg, 1b7i; Lys61Ile, 2jia have been deposited in the Protein Data Bank, Brookhaven National Laboratory, Upton, NY.

‡ Present address: Institute for Molecular Biology and Biophysics, ETH-Zurich, Zurich, Switzerland.

§ To whom correspondence should be addressed. Tel.: 613-533-6277; Fax: 613-533-2497; E-mail: jia@crystal.biochem.queensu.ca.

¹ The abbreviations used are: AFP, antifreeze protein; TH, thermal hysteresis; ASA, accessible surface area.

which in turn inhibits the growth of ice crystals. The difference between the temperature at which the ice begins to grow (burst point) and the temperature at which the ice crystal melts is known as thermal hysteresis (TH) and is used as a measure of AFP activity.

Recently, several structures of type III AFP have been determined (4–6). Based on the high-resolution x-ray structure (4), a model was proposed whereby surface adsorption occurs through a hydrogen bond match between the side chains of Gln-9, Asn-14, Thr-15, Thr-18, Gln-44, and the ice prism plane (10 $\bar{1}0$). These polar residues form part of a flat, amphipathic face that is thought to be the ice-binding surface (Fig. 1). However, the significance of the contribution from hydrogen bonds to the AFP-ice interaction has been questioned by several studies since then. A supposedly conservative change of Thr to Ser in type I AFP led to a large loss of TH activity, whereas a change to the hydrophobic residue valine, which is a better space-filling match, caused only a small loss (7, 8). In the study of the high precision NMR structures of type III AFP (5), the authors argue that hydrogen bonds alone are not sufficient to explain the affinity of the protein for ice because formation of hydrogen bonds between ice and solvent water is enthalpically more favorable because of the "perfect" alignment of water with ice. The less favorable interaction between ice and AFP could be overcome by gain in entropy because of the release of protein-associated water into the bulk solvent. Shape complementarity was examined experimentally by mutating Ala-16, a residue that is located in the center of the putative ice-binding surface (9). The loss of activity in these mutants approximately correlates with the size of the residue substituted for Ala. However, structural interpretation of these changes was complicated by shifts in adjacent residues because of the tight packing of residues at the surface.

A fundamental problem in testing binding hypotheses is the difficulty in analyzing interactions in a quantitative manner. In the crystallographic structure determination of the SP-isoform of type III AFP (6), it is also argued that hydrogen bonds are insufficient for tight binding and that flatness is perhaps a more important factor. A flatness search algorithm was designed and used to show that the proposed ice-binding surface is the flattest in the SP-isoform of type III AFP. However, in the case of the QAE-isoform of type III AFP, the putative ice-binding surface is only the second flattest plane,² which questions the relative significance of flatness.

In this study, we prepared and expressed an additional 10 mutants of the eel pout type III AFP and determined their structures to probe their contribution to the activity of the protein. Data from previous studies were also re-examined and compared. In addition, we have used a neural network (10, 11) to predict TH activity. Analysis of the structures and the neural

² Brent Wathen, personal communication.

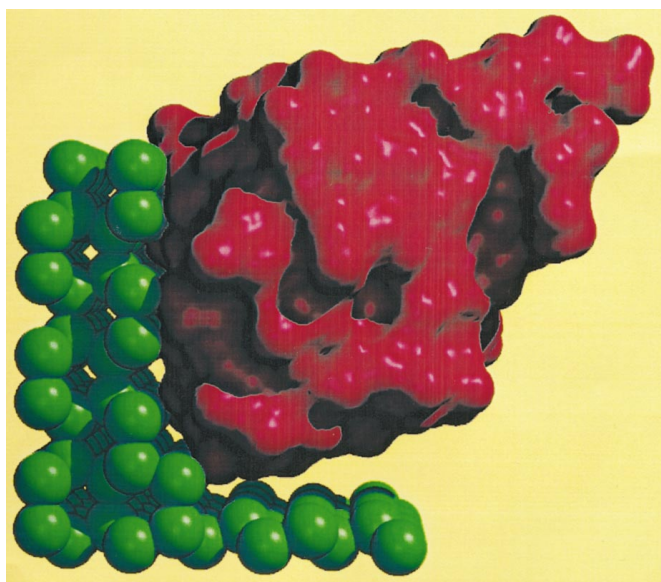


FIG. 1. Solvent-accessible molecular surface of type III AFP (red) docked to a Corey-Pauling-Koltun (CPK) representation of ice (green).

network results shows that changes in van der Waal's interactions and to a lesser extent, hydrogen bonds, are responsible for the loss of activity in type III AFP mutants.

EXPERIMENTAL PROCEDURES

Activity Measurement and X-ray Crystallography of Type III AFP—Mutants of the type III AFP QAE-isoform were made by site-directed mutagenesis as described by Chao *et al.* (12). Thermal hysteresis activity of the mutant proteins was measured and expressed as a percentage of wild-type activity. The mutant proteins crystallized under similar conditions to the wild type proteins (13) with slight variations in ammonium sulfate concentration and pH. Diffraction data (Table I) were collected using a MAR Research imaging plate equipped with a Rigaku rotating anode generator. The data were processed using DENZO/SCALEPACK (14), and structures were refined using X-PLOR (15). The mutated side chain was substituted with Ala in the model in the first round of refinement to prevent bias in the structure determination. In the second round of refinement, Ala was replaced with the mutated residue. The final refinement statistics for the mutant structures are shown in Table 2.

Interpretation of X-ray Structures—Hydrogen bond donors and acceptors involved in ice binding were defined as those oxygen and nitrogen atoms within 2.4–3.5 Å of the modeled ice layer oxygen atoms. The overlap of the mutant protein structures with the native structure was performed using LSQKAB from the CCP4 program (16). Only main-chain atoms were included in the overlap. The positional error of the structures was determined using Luzzati plots (17). Water molecules in the mutant structures were defined as conserved if they were located within 0.45 Å of the water molecule in the wild-type structure, that is, the 0.20-Å positional error from the wild-type protein plus 0.25 Å (mean positional error of the mutants). The figures were generated using Molscript (18) and Raster3D (19).

Neural Network Analysis—As an alternative and quantitative approach to visual analysis, a feed-forward neural network was constructed and tested using the Stuttgart neural network system (SNNS) (20). The SNNS package was chosen for creating and executing the neural network because of its ease of use, availability, and flexibility. Twelve properties (Table III) determined using VADAR (21) of the 16 mutants (9 mutants from this study and 7 previously made ones) and wild-type protein were used as input data, which were scaled such that the lowest value was 0, and the highest was 1. To simplify the analysis and interpretation, only proteins with single mutations were included. Results at the output node represent TH activity scaled from 0 to 1, where 0 is the lowest activity, and 1 is 100% wild-type activity. Twelve input nodes resulted in overtraining, where the neural network was able to predict precisely the TH activity of mutants given in training but not those that were left out. To reduce the number of input nodes, which reduces overtraining, principal component analysis was performed using the NCSS statistics package (22). The first component is a linear

combination of all scaled properties; the second and higher cardinal components cover variance that was not explained in the first component and are orthogonal to all other components. Six components were used to construct a network containing 6 input nodes, 20 hidden nodes, and 1 output node. The optimum number of hidden nodes was empirically determined by varying this number from 8 to 40 in increments of 4. The initial weights of connections between nodes were randomly set to values between -1 and 1 , whereas weights after training varied from -10 to 10 . Network training was performed for 5000 or more cycles, and default values were used for the remaining parameters. The ability of the network to predict TH activity was tested by cross-validation, which consisted of leaving the TH activity and six components of one mutant out before training. Then the properties of the mutant left out were applied to the trained neural network, and the TH activity was predicted. This was repeated for each of the 17 structures (wild-type protein and 16 mutants). The correlation coefficient established by the "leave-one-out" cross-validation procedure (23) between the predicted and real activity values was calculated using NCSS (22) and used as an indicator of the predictive power of the neural network. To determine which properties were responsible for the predictive ability of the neural network, each group of properties was left out in turn (Table IV), and the cross-validation procedure was repeated. Changes in the correlation coefficient were used as an indicator of the importance of the group of properties in predicting TH activity and, therefore, in the protein-ice interaction.

RESULTS AND DISCUSSION

Validation of X-ray Data—The structure determination and refinement statistics of the 11 mutant structures determined in this study are shown in Tables I and II. With the exception of N- and C-terminal residues and Val-15 in the Thr15Val mutant, there was no ambiguity in side-chain positions. A composite Ramachandran plot of all newly determined structures shows that most non-glycine residues were in the most favored region, a few were in the allowed regions, and none in the disallowed regions (Fig. 2).

Selection of Residues for Mutagenesis—We have combined all the structure information for the QAE-isoform type III AFP mutations previously and newly made (Table V) to provide a comprehensive understanding of how the protein could bind to ice and prevent its growth. To clarify the roles of the various residues, the mutations were based on residues that are proposed to hydrogen bond to ice (Gln-9, Asn-14, Thr-15, Thr-18, and Gln-44), residues that surround these and may interact with the ice through interactions other than hydrogen bonds (Ala-16, Val-20, and Met-21), and those not located on the putative ice-binding face (Ser-24, Pro-29, Pro-33, Glu-35, Ser-42, Arg-39, Asn-46, Arg-47, Asp-58, and Lys-61). The mutations are grouped according to the three regions shown in Fig. 3: (a) residues located in the "top" part of the putative ice-binding plane; (b) residues located in the "bottom" part of the proposed ice-binding plane; (c) residues located along the "bottom" of the protein; (d) residues mainly located away from these regions. Although assigning orientations and region boundaries to the protein may be seen as arbitrary, it simplifies the presentation and explanation of the interaction between type III AFP and ice.

Mutation of Residues in Region A—As shown in Table I, mutation of residues in region A resulted in some loss of TH activity, most with TH activities at 70% that of wild-type AFP. Mutation of Thr-18 to Ser resulted in no loss of activity, suggesting that a hydrogen bond between Thr-18O_γ and ice is the important factor in the interaction. For the Thr18Ser mutant, the position of the Ser-18O_γ did not change as compared with that of Thr-18O_γ. However, mutation of Thr-15 to Ser decreased TH activity by 30%, even though Ser-15O_γ would still be able to hydrogen bond to the modeled ice. It is possible that the Ser-15 hydroxyl group, unlike that of Thr-15, is no longer in an ideal position to hydrogen bond to the ice. This is supported by the x-ray structure, where the Ser-15O_γ was translated 1 Å

TABLE I
X-ray reflection data statistics

The data for the wild-type protein has been published previously (4).

Protein	Resolution	Total reflections	Unique reflections	Completeness	R_{sym}
	Å			%	
Wild type	1.25	46,873	16,529	96.6	0.049
Asn14Gln	2.0	22,376	5,329	91.0	0.070
Gln9Thr/Gln44Thr	2.3	13,186	2,599	91.0	0.082
Thr15Ser	1.75	21,851	6,260	98.0	0.079
Thr15Ala	1.7	20,580	6,915	99.4	0.064
Thr15Val	2.3	7,150	2,702	93.6	0.091
Thr18Ser	1.65	40,730	7,566	99.6	0.066
Thr18Ala	2.05	11,371	3,926	96.1	0.079
Val20Ala	1.65	27,744	7,106	93.3	0.065
Met21Ala	2.1	11,961	3,554	94.7	0.075
Arg47His	2.5	7,172	2,100	90.8	0.063

TABLE II
Structure refinement statistics

The data for the wild-type protein has been published previously (4).

Data set	Wild type	Asn14Gln	Gln9Thr/ Gln44Thr	Thr15Ser	Thr15Ala	Thr15Val	Thr18Ser	Thr18Ala	Val20Ala	Met21Ala	Arg47His
Unit cell a (Å)	32.33	33.06	32.64	33.28	33.31	32.71	32.72	32.71	33.17	32.59	32.44
b (Å)	38.98	44.07	38.94	39.88	39.90	39.38	39.13	39.37	40.06	39.53	38.71
c (Å)	45.48	44.82	46.38	44.67	46.38	46.51	46.62	46.51	47.45	45.93	47.25
Resolution range (Å)	8.0–1.25	8.0–2.0	8.0–2.3	8.0–1.75	8.0–1.7	8.0–2.3	8.0–1.65	8.0–2.0	8.0–1.65	8.0–2.1	8.0–2.5
R_{factor} (%)	19.8	18.5	18.3	18.3	18.5	16.3	19.0	17.2	18.8	16.6	15.5
(R_{free} , 5%)	22.5	26.1	29.2	26.2	23.0	22.3	21.3	23.8	20.8	22.7	25.1
No. of water molecules	90	48	41	58	46	42	79	59	51	72	42

TABLE III

TH activity measured and predicted by the neural network

TH activity is expressed as percent of wild-type protein at 1 mg/ml. Predicted TH activity was determined with the neural network after the properties for that protein were left out of training.

Protein	Measured TH	Predicted TH
	% wild type	% wild type
Wild type	100	63
Ala16Cys	100	99
Thr18Ser	100	96
Arg47His	100	100
Ala16Met	85	89
Val20Ala	80	92
Met21Ala	80	85
Ala16Thr	75	46
Thr15Ala	70	92
Thr15Ser	70	85
Thr18Ala	70	24
Asn14Gln	67	70
Ala16Arg	60	35
Thr15Val	54	31
Ala16Tyr	33	22
Ala16His	25	77
Thr18Asn	10	50

and rotated approximately halfway between the methyl and hydroxyl groups of the wild-type Thr-15 (data not shown). In addition, ice crystals in the presence of Thr15Ser were distinguishable from those of the wild type. They did not grow until just before the burst point, at which time hollow spicules began to form at various points along the crystal. In contrast to the Thr18Ser mutation, the methyl groups of Thr-15 may be required to correctly orient the hydroxyl group so that it is able to effectively hydrogen bond to ice. It would then follow that the 30% loss in activity in the Thr15Ala and Thr18Ala mutants occurred for similar reasons, principally the loss of a hydrogen bond between the protein and ice. The Thr15Val mutation shows that the substitution of the hydroxyl group for a methyl group resulted in approximately half of the TH activity. The electron density of this side chain is poorly defined, unlike the side chains in the other mutant structures. Val-15 may there-

TABLE IV

Four groups with 12 properties used in the neural network analysis and correlation coefficients after the leave-one-out cross-validation procedure

Neural network training and the cross-validation procedure are described under "Experimental Procedures."

Group	Correlation coefficient after cross-validation with group removed
All properties	0.60
Protein dimensions	0.29
Total volume	
Total ASA	
ASA of side chains	
Hydrophobic	-0.04
Fractional non-polar ASA	
Percent side-chain hydrophobic ASA	
ASA of carbon	
Polar	0.53
ASA of oxygen	
ASA of nitrogen	
Fractional polar ASA	
Charged	0.33
ASA of charged nitrogen	
ASA of charged oxygen	
Fractional charged ASA	

fore not have a well defined position and may exist as one of two main orientations. In one, the side chain can be found in a similar orientation to the Thr-15; in the second orientation, the methyl group pointed toward the hydrophobic core of the protein. When in the latter position, the distance between Val15C₂ and the nearest modeled ice oxygen is 1.85 Å. The resulting steric clash may explain why the presence of valine at this position had a greater effect than the mutation to alanine. It would be of interest to know whether Thr18Val behaves similarly, but several attempts at refolding expressed Thr18Val have failed.

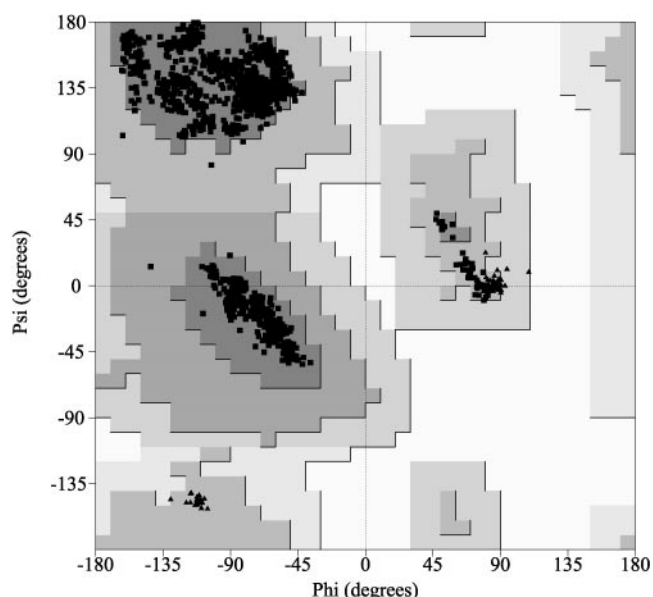


FIG. 2. **Composite Ramachandran plot of type III AFP mutants.** The data were generated using PROCHECK (16). The dark gray area represents most allowed regions, whereas the medium gray areas represent allowed regions. Glycine residues are represented by triangles with other residues represented by squares.

TABLE V

Mutation of type III AFP residues and the structure and TH activity

Regions A–D are defined in Fig. 3. Structures were determined as described under “Experimental Procedures.” TH activity is expressed as percent of wild-type protein at 1 mg/ml. The structures of wild-type protein, Thr18Asn, and Asn14Ser/Gln44Thr were determined previously (4). ND, not determined.

Region	Protein	Structure resolution	TH
		Å	% wild type
A	Wild type	1.25	100
	Gln9Thr	ND	71
	Thr15Ala	1.7	70
	Thr15Ser	1.75	70
	Thr15Val	2.3	54
	Thr18Ala	2.0	70
	Thr18Asn	2.0	10
	Thr18Ser	1.65	100
	Val20Ala	1.65	80
	Met21Ala	2.1	80
B	Asn14Gln	2.0	67
	Asn14Ser	ND	25
	Gln44Thr	ND	47
	Asn14Ser/Gln44Thr	2.3	10
A/B	Gln9Thr/Gln44Thr	2.6	ND
C	Arg47His	2.5	100
	Asp58Ser	ND	100
D	Lys61Ile	1.6	100
	Ser24His	ND	100
	Glu35Lys	1.65	100
	Ser42Gly	1.8	100
	Asn46Ser	1.9	100

The mutant Gln9Thr has approximately the same loss of TH activity as Thr15Ala and Thr18Ala. The structure of Gln9Thr has not been determined, but that of the double mutation Gln9Thr/Gln44Thr has. Because these residues are far apart in the structure, it is assumed that the mutation at Gln-44 does not have an effect on the position of Gln-9. An examination of Gln9Thr in the double mutant showed that Thr-9 is too far away (4.5 Å) from the modeled ice to form a hydrogen bond.

Val-20 and Met-21 are hydrophobic residues located in the proposed ice-binding region. No major changes in the protein structure were seen in either the Val20Ala or Met21Ala mutants. In Met21Ala, there were two minor shifts of Thr-180_γ

and Gln9N_ε toward each other, because the methionine side chain no longer separates the two, but these would not drastically affect the ability of Thr-18 and Gln-9 to form hydrogen bonds with the modeled ice surface. These residues appear to contribute to the relative flatness of the proposed ice-binding surface, which could allow favorable weak interactions, such as van der Waal’s interactions, to occur. When Val-20 or Met-21 are mutated to Ala, the previously flat surface becomes recessed. The mutations are tolerated fairly well, with only a 20% loss in TH activity. The opposite case, where a substitution causes the side chain to extend above the flat surface, can cause a drastic decrease in TH activity. One example is Thr18Asn, where 90% of the activity was lost (4). The longer side chain may prevent the other ice-binding residues from interacting at the same time with the ice surface. A similar effect is seen with the Ala-16 mutants, where small steric additions to the ice-binding surface by Cys, Met, Thr, or Val resulted in a small decrease in TH activity, whereas bulky groups such as His or Tyr caused a large decrease (9).

Mutation of Residues in Region B—In a previous AFP-ice binding model, it was proposed that Asn-14 and Gln-44 are the first residues to bind to ice (4). Mutation of either of these residues to one with a shorter side chain that still had the ability to form hydrogen bonds, although not necessarily to ice, resulted in a large loss of activity. In the structure of the double mutant Asn14Ser/Gln44Thr, the Ser-14O_γ forms a hydrogen bond with Lys61N_ε, so that Ser-14 can no longer effectively hydrogen bond to modeled ice. The potential loss of the hydrogen bond at residue Asn-14 is drastic, because the single mutant Asn14Ser has only 25% TH activity. In the case of Asn14Gln, the longer side chain pushes Lys-61 away to make more space so that Gln14O_ε is still able to hydrogen bond with Lys61N_ε. However, in the x-ray structure, Gln14N_ε is now too far from the modeled ice to form a hydrogen bond. The decrease in TH activity is not as great as with Asn14Ser (67% versus 25% TH activity). Conceivably, Gln-14 may be able to alternate between hydrogen bonding to Lys-61 and to the ice. A mutation at Gln-44 to threonine also results in a structure that is no longer able to hydrogen bond to modeled ice. The activity loss is less severe (50%). The multistep process, where Asn-14 binds first before other hydrogen residues in the ice face, is still a possibility because if other residues bound before Asn-14 did, one might expect mutations of them to cause a similar or more severe decrease in TH activity.

Mutation of Residues in Regions C and D—Mutations in these two regions were made to explore the potential existence of additional ice-binding surfaces on type III AFP. In the case of Pro29Ala and Pro33Ala mutations, the loss of 50% activity was probably the result of changes in the protein backbone conformation because of the structural role often played by proline residues (24). Aside from these two exceptions, mutation of other residues in regions C and D resulted in no detectable loss of TH activity. These mutants are, however, subject to a distinction between those that allowed ice crystal growth during the measurement of TH and those that did not. Residues in region D resulted in no loss of activity and no change in ice crystal morphology. Therefore residues Ser-24, Glu-35, Arg-39, Ser-42, and Asn-46 are not involved, directly or indirectly, in the interaction with ice, and there are no additional ice-binding planes along the top and sides of the protein. In region C, mutation of residues Arg-47, Asp-58, or Lys-61 resulted in a protein with a burst point similar to or slightly higher than that of wild-type protein. Typically, there was slow growth of ice, and there was more variation in TH values. In the case of Lys-61, there could be an interaction between the side chain and the basal plane of ice. The potential formation of additional

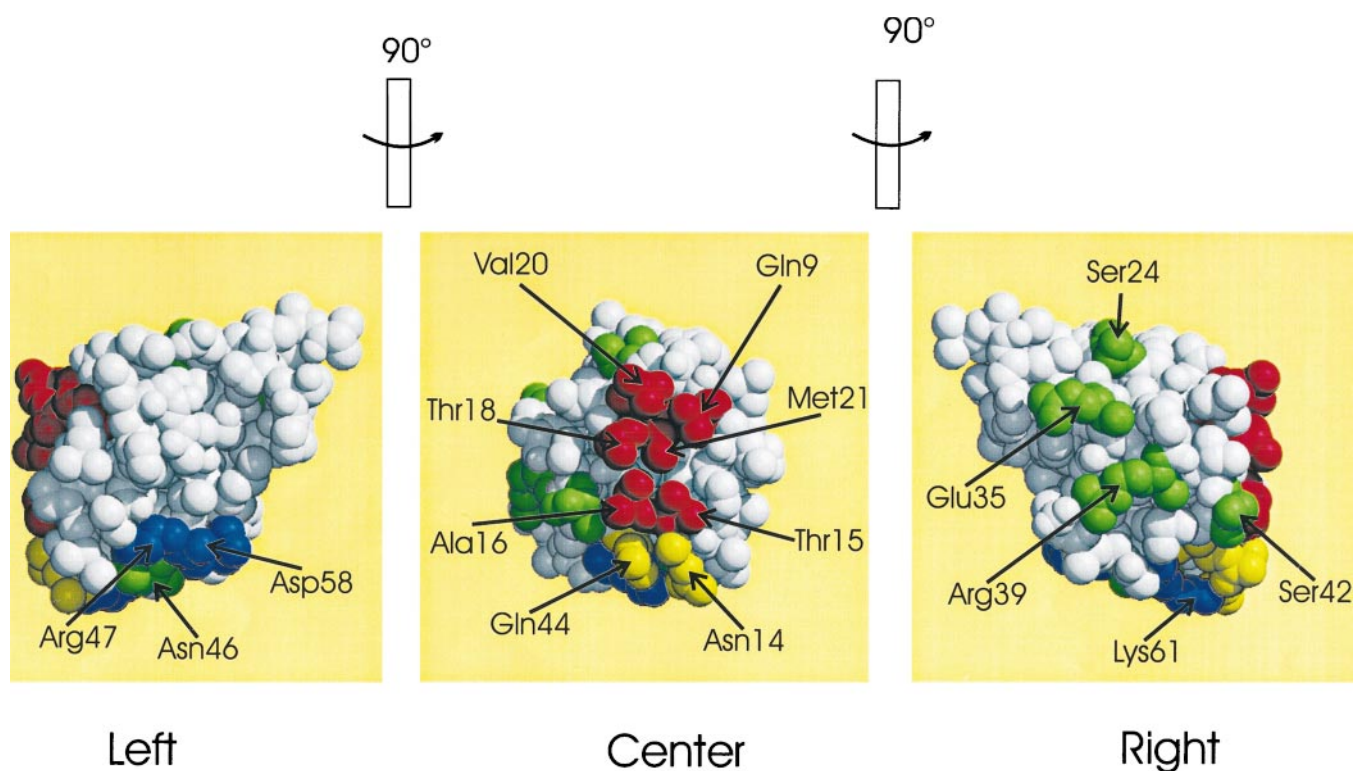


FIG. 3. Location of mutations in type III AFP; three views of a Corey-Pauling-Koltun (CPK) model of the wild-type III AFP with changed residues marked in color. Red, region A; yellow, region B; blue, region C; green, region D.

hydrogen bonds between the lower protein surface and the basal plane would add to the binding energy, increasing the strength of the interaction. Although residues Arg-47 and Asp-58 are not located on the bottom of AFP, they may indirectly affect the ability of Lys-61 to bind to the basal plane of ice itself or to hydrogen bond to Asn-14. For the latter case, this suggests that residues in region C may not bind to the modeled ice itself, but that losses in activity could be because of the inability of Lys-61 to correctly position Asn-14 for binding to ice.

Neural Network Training—The basic problem of modeling an AFP binding to ice is that no methods that involve the direct detection of interactions between protein and ice (such as solid state NMR) have been reported; thus, the analysis of structure/function relationships have been done in an indirect, qualitative manner. Even more significantly, the modeling of AFP to ice is speculative. Therefore, the TH activities and structures determined in this study and previously were used in a neural network to achieve a quantitative analysis of this interaction that involves proteins but not any AFP-ice model.

Properties were chosen and grouped to account for four possibly important interactions of protein-ice binding. The first group, protein dimensions, consists of total volume, total accessible surface area (ASA), and total ASA of the side chains, where ASA is defined as the area of the protein surface that is in contact with solvent (21). The second group, hydrophobic properties, consists of the nonpolar fractional ASA, percent side-chain ASA, and ASA of carbon. The third group, polar properties, consists of ASA of oxygen, ASA of nitrogen, and fractional ASA of polar atoms, whereas the fourth group, charged properties, consists of ASA of charged nitrogen, ASA of charged oxygen, and fractional ASA of charged groups.

To run the neural network, it is first necessary to optimize several training parameters. Overtraining, that is, the inability of a neural network to predict data it has not seen before, although predicting given data well, increases with the number

of input nodes used. To reduce the number of input nodes, principal component analysis was performed on the 12 properties (25). Six components explained 85% of the variance of all of the properties, whereas fewer or more components resulted in a neural network that could not be trained. Therefore, the six components were used for the six input nodes in the neural network. The size of the hidden layer was varied from 8 to 40 nodes by increments of 4 until the highest correlation factor between experimental and predicted TH activity was determined, which resulted in 20 nodes. The number of training cycles used was 5000 or more cycles.

Neural Network Analysis of Mutants—With the chosen parameters, the average percent difference between the experimentally determined TH activity and predicted TH activity of structures included in training was <0.1%, demonstrating that the network was well optimized.

The correlation between experimental and predicted TH activity for structures used in the leave-one-out cross-validation was 0.60, indicating that the neural network could successfully predict the activity of mutants left out. This value is significant because of the leave-one-out cross-validation procedure (23). Table III shows the measured and predicted activity of wild-type type III AFP and the 16 mutants. Fourteen of the proteins had predicted TH activities that were within 50% of the measured values, with 8 of these being within 25%. Of the remaining 3 mutants, Thr18Ala, Thr18Asn, and Ala16His, the predicted activities did not correspond well with measured activities. It is not clear why these mutants failed, because similar mutations with other residues did not result in a similar failure. Overall, the neural network was well able to predict the activity of a number of diverse mutations of type III AFP. The next step was to remove each group of properties in turn, examine the network for loss of predictive power, and use this as an indicator of which properties are potentially important in type III AFP binding to ice. A simple examination of the neural network to determine the relative weight of each property was not possible

because of the principal component analysis performed and the 20 hidden nodes used.

The changes in the correlation coefficient after leaving each group out in turn before repeating the cross-validation procedure is shown in Table IV. After leaving out protein dimensions and polar or charged properties, the correlation coefficients dropped significantly, but the remaining terms were still able to give some predictive power to the neural network. Thus, these properties could also have a role in ice binding. However, when the hydrophobic group properties were removed, the correlation coefficient essentially dropped to zero. This strongly suggests that the main ability of the neural network to predict the TH activity of the mutants comes from learning about changes in hydrophobic character of the protein surface. The actual interaction could come from attractive van der Waal's interactions and require that type III AFP have a surface that is complementary to that of the ice. This is supported by the data of leaving protein dimensions out, which resulted in the second largest decrease in the correlation coefficient. These properties have no chemical basis but instead are a reflection of changes in the protein shape.

The neural network did not have information of the location of any residue. This would mean, for example, that two mutations that created a protein with identical global surface properties could not be distinguished, even though one may be located in the ice-binding face, but another is far away. Therefore, the neural network would not be able to effectively determine the importance of hydrogen bonds in the interaction. This is not a flaw in the design but was done so as not to bias the neural network with modeled ice. Combining results from structural analysis and neural network, we envision a mechanism where 1) hydrogen bonds could search for and recognize the correct ice-binding plane and 2) attractive interactions are facilitated by shape complementarity, which allows the protein to remain bound to ice.

Other Factors in Ice Binding—Two additional issues we examined could not be easily integrated into the neural network. Given the large number of mutant structures, the conservation of positioned water molecules was also examined. Many clusters of water molecules were found in the N- and C-terminal region of the protein, which is located away from the putative ice-binding face, whereas few clusters were in the ice-binding region (data not shown). In addition, the structures of three preparations of wild-type proteins have been determined independently, one of which was crystallized at 4 °C (3). None of these structures had water clusters resembling the modeled ice. One may speculate that the water around the ice-binding surface is kept mobile to decrease the energy required to remove them before binding to ice. Such an arrangement may also help to prevent type III AFP acting as an ice seed, because water molecules aligned in an ice-like fashion could promote ice crystal formation before the protein could bind to an existing ice crystal.

Analysis of the structures shows one region where no positioned water molecules are present. Residues in this region are surface-exposed hydrophobic groups (Leu-10, Ile-13, Leu-19, Ile-37, Val-41, and Leu-51) and have no water molecules within 4 Å of any of the atoms in the side chain. They form a hydrophobic ring that circumscribes approximately 2/3 around the protein and is located behind the proposed ice-binding face. The ring is broken by Lys-61, which has a cluster of conserved waters. Ongoing studies of mutation of the residues in this ring to alanine show a dramatic loss of activity.³ Therefore, part of type III AFP stopping of ice growth may come from the ability of the protein to exclude water from near the ice surface or to keep this water in such a state so that it cannot bind to the ice surface. It is, however, difficult to quantitate the contribution of the hydrophobic ring to the overall interaction and testing of this hypothesis requires further experimentation.

Acknowledgments—We thank Sherry Gauthier and Dr. Qilu Ye for their excellent technical assistance. We are also grateful to Brent Wathen for help running the neural network.

REFERENCES

- Davies, P. L., and Sykes, B. D. (1997) *Curr. Opin. Struct. Biol.* **7**, 828–834
- Raymond, J. A., and DeVries, A. L. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 2589–2593
- Ye, Q., Leinala, E. K., and Jia, Z. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54**, 700–702
- Jia, Z., DeLuca, C. I., Chao, H., and Davies P. L. (1996) *Nature* **384**, 285–288
- Soennichsen, F. D., DeLuca, C. I., Davies, P. L., and Sykes, B. D. (1996) *Structure (Lond.)* **4**, 1325–1337
- Yang, D. S., Hon, W. C., Bubanko, S., Xue, Y., Seetharaman, J., Hew, C. L., and Sigheri, F. (1998) *Biophys. J.* **74**, 2142–2151
- Chao H., Houston, M. E., Jr., Hodges, R. S., Kay, C. M., Sykes, B. D., Loewen, M. C., Davies, P. L., and Sonnichsen, F. D. (1997) *Biochemistry* **36**, 14652–14660
- Haymet, A. D., Ward, L. G., Harding, M. M., and Knight, C. A. (1998) *FEBS Lett.* **430**, 301–306
- DeLuca, C. I., Davies, P. L., Ye, Q., and Jia, Z. (1998) *J. Mol. Biol.* **275**, 515–525
- Presnell, S. R., and Cohen, F. E. (1993) *Annu. Rev. Biophys. Biomol. Struct.* **22**, 283–298
- Müller, B., Reinhardt, J., and Strickland, M. T. (1995) *Neural Networks. An Introduction*, 2nd Ed., Springer-Verlag New York Inc., New York
- Chao, H., Soennichsen, F. D., DeLuca, C. I., Sykes, B. D., and Davies, P. L. (1994) *Protein Sci.* **3**, 1760–1769
- Jia, Z., DeLuca, C. I., and Davies, P. L. (1995) *Protein Sci.* **4**, 1236–1238
- Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307–326
- Brunger, A. T. (1992) *X-PLOR Manual Version 3.1. A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT
- Collaborative Computing Project No. 4 (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 760–763
- Luzzati, V. (1952) *Acta Cryst.* **5**, 802–810
- Kraulis, P. J. (1991) *J. Appl. Cryst.* **24**, 946–950
- Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol.* **277**, 505–524
- Zell, A., Mache, N., Sommer, T., and Korb, T. (1991) *Informatik-Fachberichte* **285**, 254–263
- Wishart, D. S., Willard, L., Richards, F. M., and Sykes, B. D. (1994) *VADAR: A Comprehensive Program for Protein Structure Evaluation*, Version 1.2. University of Alberta, Edmonton, Alberta, Canada
- Hintze, J. (1998) *NCSS Statistical Package*, NCSS Kaysville, Utah
- Brunger, A. T. (1995) *Nature* **355**, 472–474
- Chao, H., Davies, P. L., Sykes, B. D., and Sonnichsen, F. D. (1993) *Protein Sci.* **2**, 1411–1428
- Hotelling, H. (1933) *J. Ed. Psych.* **24**, 417–441, 498–520

³ J. Baardsnes, and P. L. Davies, unpublished data.