Published in final edited form as:
J Am Chem Soc. 2008 July 30; 130(30): 9695–9701. doi:10.1021/ja8013538.

X-ray Structure of Snow Flea Antifreeze Protein Determined by Racemic Crystallization of Synthetic Protein Enantiomers

Brad L. Pentelute[†,‡], **Zachary P. Gates**[†,‡], **Valentina Tereshko**[†,§], **Jennifer L. Dashnau**[⊥], **Jane M. Vanderkooi**[⊥], **Anthony A. Kossiakoff**[†,§], and **Stephen B. H. Kent**[*,†,‡,§]

[†]Institute for Biophysical Dynamics, University of Chicago, Chicago, Illinois 60637

[‡]Department of Chemistry, University of Chicago, Chicago, Illinois 60637

[§]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637

[⊥]Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Abstract

Chemical protein synthesis and racemic protein crystallization were used to determine the X-ray structure of the snow flea antifreeze protein (sfAFP). Crystal formation from a racemic solution containing equal amounts of the chemically synthesized proteins D-sfAFP and L-sfAFP occurred much more readily than for L-sfAFP alone. More facile crystal formation also occurred from a quasi-racemic mixture of D-sfAFP and L-Se-sfAFP, a chemical protein analogue that contains an additional -SeCH2- moiety at one residue and thus differs slightly from the true enantiomer. Multiple wavelength anomalous dispersion (MAD) phasing from quasi-racemate crystals was then used to determine the X-ray structure of the sfAFP protein molecule. The resulting model was used to solve by molecular replacement the X-ray structure of L-sfAFP to a resolution of 0.98 A. The L-sfAFP molecule is made up of six antiparallel left-handed PPII helixes, stacked in two sets of three, to form a compact brick-like structure with one hydrophilic face and one hydrophobic face. This is a novel experimental protein structure and closely resembles a structural model proposed for sfAFP. These results illustrate the utility of total chemical synthesis combined with racemic crystallization and X-ray crystallography for determining the unknown structure of a protein.

Introduction

Antifreeze proteins are a structurally diverse class of macromolecules that have in common the abilities to effect thermal hysteresis of water (nonequilibrium freezing point depression) and to inhibit ice recrystallization (growth of large ice crystals at the expense of small ones).[1] These proteins are critical to the existence of life at subzero temperatures[2,3] and also have potential practical applications, e.g. for the storage of organs for use in transplantation. Both antifreeze proteins and antifreeze peptides function by binding to the surface of ice crystals and preventing further crystal growth.[4] The exact molecular basis of the binding of antifreeze proteins to ice crystals is still subject to investigation;[5] it was originally thought that antifreeze peptides bind to the surface of an ice crystal through H-bonding of polar side chain moieties, for example in Thr or Lys residues, suitably disposed on the peptide molecule. More recent studies of antifreeze proteins have pointed toward key roles for shape complementarity between a flat hydrophobic surface of the protein and the surface of the ice crystal, with both van der

E-mail: E-mail: skent@uchicago.edu.

Waals-type forces and H-bonds contributing to keep the hydrophobic protein face bound tightly to the ice crystal.[1]

Two years ago, Davies and co-workers reported the discovery of a novel antifreeze protein from the Canadian snow flea.[6] The snow flea antifreeze protein (sfAFP) studied here is the smaller of two protein isoforms that they isolated by ice-affinity purification of snow flea homogenates. The predicted 81-residue amino acid sequence of sfAFP is glycine rich, and is characterized by a repetitive Gly-Xaa-Xaa motif, the presence of several proline residues, and by two intramolecular disulfides, the connectivity of which has not been experimentally determined (Scheme 1). sfAFP has no known homologues, antifreeze protein or otherwise, and no experimental structure has been reported for the folded protein molecule. Recently, Davies and coworkers have proposed a structural model for sfAFP.[7] Experimental structure determination has been precluded in the first instance by a simple lack of material—isolation of the natural product gives only microgram quantities, and recombinant expression is reportedly difficult and does not yield significant amounts of protein.[7]

Chemical protein synthesis, based on modern ligation methods,[8,9] provides a route to the preparation of protein molecules up to 200 or more amino acid residues in size.[10] Recent refinements have made total chemical synthesis an increasingly practical way of obtaining useful amounts of protein for study. [11–14] In an accompanying paper, we report the total chemical synthesis of the mirror image forms of the snow flea antifreeze protein: L-sfAFP and D-sfAFP;[15] the antifreeze activities of both of these folded, synthetic proteins were verified by ice recrystallization inhibition assays.[15] With a ready supply of sfAFP to hand, we set out to determine the experimental structure of sfAFP using racemic protein crystallography, a technique that is made possible only by using chemical protein synthesis.

While X-ray structures have been determined for many thousands of distinct protein molecules, it is not infrequently observed that a particular protein may be difficult to crystallize or may form crystals unsuitable for the collection of single-crystal X-ray diffraction data. This has become one of the principal obstacles to the high throughput determination of protein X-ray structures. As will be described below, we encountered severe difficulties in the crystallization of sfAFP, which we were able to overcome as follows.

Proteins found in nature (i.e., proteins made up of L-amino acids and glycine) are inherently chiral entities and can therefore crystallize only in one of the 65 “biological” space groups that lack an inversion center or mirror plane, out of the total of 230 unique space groups that describe all possible crystal symmetries.[16] The frequency of protein crystallization in different space groups is markedly nonuniform, with most protein crystals belonging to one of just a few space groups. $P2_12_12_1$ is by far the most commonly observed space group for globular protein molecules. Yeates has explained this preference for certain space groups based on an entropic model: “the favored space groups are simply less restrictive than others… they allow the molecules more rigid-body degrees of freedom and can therefore be realized in a greater number of ways.”[16] This same model predicts that, for globular proteins, a racemic protein mixture will crystallize more readily. This is so because the space group most favored for globular proteins on theoretical grounds, $P\bar{1}$, contains inversion symmetry and is thus not accessible to natural L-proteins.[16,17] However, the $P\bar{1}$ space group would in principle be accessible to the achiral entity made up of a D-protein and the corresponding L-protein as a molecular pair. The more facile crystallization of racemic protein mixtures has never been experimentally verified. Only two proteins have been crystallized as racemates; in both cases centrosymmetric racemic crystals in space group $P\bar{1}$ were obtained.[18,19] These pioneering studies showed that it is possible to obtain centrosymmetric crystals from a racemic protein mixture; however, in neither case was crystallization of the natural L-protein limiting, so the

studies did not speak to whether or not racemic crystallization could be used to facilitate obtaining crystals from a difficult-to-crystallize protein target.

In this paper we report the use of racemic crystallization[16–22] to obtain high-quality crystals of the protein sfAFP for X-ray diffraction studies. Crystal formation occurred much more readily from a racemic mixture of D-sfAFP and L-sfAFP than with the L-protein alone. More facile crystal formation also occurred from a quasi-racemic mixture of D-sfAFP and L-Se-sfAFP, a chemical protein analogue that contains an additional -$SeCH_2$- moiety at one residue and thus differs slightly from the true enantiomer. Multiple wavelength anomalous dispersion (MAD) phasing of diffraction data from a quasi-racemate crystal was then used to solve the X-ray structure of the sfAFP protein molecule. The resulting model was used to solve by molecular replacement the structure of L-sfAFP to a resolution of 0.98 Å. The X-ray structure of the true racemate made up of the protein enantiomers L-sfAFP and D-sfAFP in space group $P\bar{1}$ was also analyzed.

Results and Discussion

Crystallization of L-sfAFP

Using sfAFP prepared by total chemical synthesis,[15] extensive crystallization trials were carried out over a period of six months with L-sfAFP alone, varying temperature, protein concentration, and the concentration of $(NH_4)_2SO_4$ or other precipitants. Crystals were eventually obtained, with difficulty. By screening selected sections of ~30 crystals, grown at room temperature (~23 °C) by vapor diffusion, we were able to collect a synchrotron diffraction data set to 0.98 Å resolution. These data were indexed in the $P2_1$ space group and had unit cell parameters a = 16.7 Å, b = 74.3 Å, c = 17.7 Å, β = 102.2°, and had one sfAFP molecule per asymmetric unit. The solvent content for this crystal was unusually low at 25%. Exhaustive attempts to crystallize heavy atom derivatives of L-sfAFP for use in phase determination failed, and no structures of homologous protein molecules were available, so we were unable to solve the X-ray structure.

Racemic Crystallization

The total chemical synthesis of the two enantiomeric molecules, D-sfAFP and L-sfAFP, was carried out as previously described.[15] In trials using a racemic solution containing equal amounts of D-sfAFP and L-sfAFP we found that crystals were obtained in ~50% of the 194 conditions initially explored using the commercially available Hampton index, crystal screen and crystal screen II. This was in marked contrast to what we had observed in attempts to crystallize L-sfAFP. Figure 1 shows results obtained from simultaneous crystallization experiments under representative conditions for L-sfAFP alone, and for a racemic solution of D-sfAFP and L-sfAFP under the same conditions.

Diffraction data from a sampling of the racemic crystals formed in our initial screen revealed that, under all conditions examined, the protein racemate crystallized with triclinic symmetry and had unit cell parameters a = 28.6 Å, b = 32.4 Å, c = 59.8 Å, α = 88.7°, β = 89.2°, γ = 73.4° with two sfAFP molecules in the asymmetric unit. The $\langle |E^2 - 1| \rangle$ statistic calculations in program XPREP indicated the centrosymmetric space group $P\bar{1}$

Crystallization of a Quasi-Racemate for Phase Determination

The next step was to solve the structure of the sfAFP molecule. It was suggested by Berg that, because of the quantization of phases in diffraction data obtained from a centrosymmetric racemic protein crystal, direct methods should be more applicable.[22] For the sfAFP protein, our preliminary attempts to solve the structure by direct methods failed. To speed up the structure determination we chose to solve the structure using experimental phases obtained

using the multiple wavelength anomalous dispersion (MAD) method.[23] We collected synchrotron X-ray data from a selenium-containing quasiracemate in which only one enantiomer contained a selenium atom. (For centrosymmetric structures, such as a protein racemate [pair of enantiomers], Friedel's law holds even in the presence of an anomalous scattering atom, and it is not possible to measure anomalous differences.[24]) The quasi-racemate was obtained by the cocrystallization of a mixture of D-sfAFP and L-Se-sfAFP. The L-Se-sfAFP was a chemically synthesized sfAFP analogue in which Asn^{11} was replaced by seleno-Cys alkylated with bromoacetamide ("pseudo-Se-Gln"; see Experimental Section for details). The more facile production of crystals by racemic crystallization was critical for obtaining crystals —we had made extensive efforts to grow crystals of the L-Se-sfAFP protein alone, with no success. By contrast, a quasiracemic solution containing equal amounts of D-sfAFP and L-Se-sfAFP crystallized readily under a variety of conditions.

Synchrotron X-ray diffraction data to a resolution of 1.2 Å were collected from crystals of the D-sfAFP and L-Se-sfAFP quasi-racemate. The quasi-racemate had unit cell parameters nearly identical to those of the true racemate. The structure was solved in space group *P*1 using MAD data collected at three wavelengths (see Experimental Section). Figure 2 shows the arrangement of D-sfAFP and L-Se-sfAFP molecules in the quasiracemate.

X-ray Structure of sfAFP

The resulting high resolution X-ray structure of sfAFP is shown in Figure 3. The protein fold observed in the X-ray structure of sfAFP is similar to the structural model recently proposed by Davies et al.[7] It consists of six left-handed PP-II helices, stacked in two antiparallel groups to form an oblong brick-shaped molecule. The turns linking the helices are in general defined by the presence of Pro residues, and the fold of the N-terminal half of the polypeptide chain is stabilized by two intramolecular disulfide bonds between residues Cys1-Cys28 and Cys13-Cys43. One face of the molecule is hydrophobic, and the other, polar/hydrophilic. The implications of this novel protein fold for antifreeze activity have been discussed.[7] It has been hypothesized that in antifreeze proteins a flat, hydrophobic face interacts snugly with the highly ordered water molecules found at the surface of an ice crystal, presenting the polar (hydrophilic) face to the surrounding solution and inhibiting further growth of the ice crystal.[25,26] Interestingly, we observed an array of highly ordered water molecules on the flat, hydrophobic face of the sfAFP protein (Figure 3). This observation is consistent with the proposed mechanism for the binding of an sfAFP molecule to an ice crystal.[7]

The molecular model from the quasi-racemate was used to solve by molecular replacement the structure of L-sfAFP using the original 0.98 Å diffraction data (Table 3). The L-sfAFP molecular structure in the $P2_1$ crystal is identical to that observed in the quasi-racemate. A portion of the $2F_o - F_c$ electron density map obtained at 0.98 Å is shown in Figure 4(a).

Diffraction data were also collected from crystals of the true racemate {D-sfAFP + L-sfAFP}. Table 4 gives the X-ray statistics for the {D-sfAFP + L-sfAFP} data set. A portion of the $2F_o - F_c$ electron density map obtained for the true racemate {D-sfAFP + L-sfAFP} at 1.0 Å is shown in Figure 4 (b). The crystals of the true racemate were refined in the $P\bar{1}$ centrosymmetric space group.

Conclusions

Crystallization of a racemic mixture of enantiomeric protein molecules prepared by total chemical synthesis has been used to experimentally determine the X-ray structure of the snow flea antifreeze protein (sfAFP). This protein was found to have a tertiary structure consisting only of stacked PPII helices, which represents a unique protein fold. The experimental structure

of sfAFP is similar to a molecular model recently proposed, and has implications for the molecular basis of the antifreeze activity of sfAFP.[7]

In this paper we have experimentally demonstrated, for the first time, that a racemic protein mixture can more readily form highly ordered crystals suitable for X-ray diffraction studies. This can greatly facilitate the X-ray structure determination of a synthetically accessible protein, as was the case for the snow flea antifreeze protein studied here. Moreover, we have shown that, at least for sfAFP, the more facile crystallization also applies to quasi-racemic mixtures in which one protein molecule contains an anomalous scattering atom for phase determination. Including the work reported here, only a handful of racemic protein crystallizations have been described. Because a mixture of enantiomers can crystallize either as a racemate or the individual enantiomers can crystallize as a conglomerate, it is too early to say whether more facile crystallization of racemic protein mixtures is a general phenomenon, as predicted by Yeates.[16] If protein racemates crystallize as a rule, then the chemical synthesis of mirror image proteins for racemic crystallization, as suggested by Yeates,[16,17] may prove to be a powerful tool in structural biology.

Experimental Section

Total Chemical Synthesis of sfAFP and Analogues

The 81-residue sfAFP polypeptide chain (Scheme 1) was synthesized by native chemical ligation of four unprotected peptides (Thz1–12-thioester; Thzl3–27-thioester; Thz28–42-thioester; Cys43–81) as described in an accompanying paper.[15] L-Se-sfAFP was prepared by incorporation of "pseudo-Se-Gln" into the Thz1–12- thioester peptide during stepwise solid-phase synthesis.[27] Boc-"[4Se]-pseudo-Gln"-OH for use in peptide synthesis was prepared by reducing L-selenocystine with sodium borohydride, alkylating at selenium with 2-bromoacetamide, and *N*-acylating with Boc_2O. The LC—MS analysis of chemically synthesized L-Se-sfAFP synthetic protein is shown in Figure 5.

Protein Crystallization

Lyophilized sfAFP was dissolved in water at 10–40 mg/mL and was used directly for crystallizations. Crystallization screening conditions were based on the commercially available Hampton index, crystal screen and crystal screen II. Crystals were grown by vapor diffusion in hanging drops at room temperature (~23 °C). The drops were generated by mixing 1 μL of protein solution with 1 μL of reservoir solution and placed against 1 mL of reservoir solution. Racemic solutions of sfAFP were prepared in water at 38 mg/mL (19 mg of D-sfAFP and 19 mg of L-sfAFP) and 19 mg/mL (9.5 mg of D-sfAFP and 9.5 mg of L-sfAFP). Simultaneous sparse matrix room temperature screens were carried out for the racemic protein solutions at both concentrations. Crystals appeared after 1 day, and after 10 days, 93 of the 194 sets of the conditions explored had crystals suitable for use in diffraction experiments. No crystals were formed from the L-sfAFP solutions using the same sets of conditions. The same approach was used for crystallization of the quasi-racemate in which crystals suitable for use in diffraction experiments appeared after 1 day in a number of conditions. Diffraction data for the quasi-racemate were collected using a crystal grown at room temperature from 0.2 M sodium malonate, pH 7, and 20% w/v polyethylene glycol 3350. Diffraction data for L-sfAFP were collected using a crystal grown at room temperature from 0.1 M bis-Tris pH = 5.5, 1.8 M ammonium sulfate. Diffraction data for the racemate was collected using a crystal grown at room temperature from 0.1 M bis-Tris pH = 6.5, 1.15 M ammonium sulfate.

Diffraction Data Collection for L-sfAFP

For low temperature data collection, selected crystals of L-sfAFP were transferred to the cryoprotectant (reservoir solution plus 20% (v/v) glycerol) for a few seconds and flash-frozen

in liquid nitrogen. Synchrotron X-ray data were collected at a temperature of 100 K at the Advanced Photon Source, Argonne National Laboratory, beamline 19BM equipped with a SBC2 detector. Images were processed and scaled with HKL2000.[28] Data were collected to a resolution of 0.98 Å (Table 3).

Diffraction Data Collection for the Racemate

For low temperature data collection, selected crystals were transferred to the cryoprotectant (reservoir solution plus 20% (v/v) glycerol) for a few seconds and flash-frozen in liquid nitrogen. X-ray data were collected at 100 K at the Advanced Photon Source, Argonne National Laboratory, beamline 23ID equipped with a MARCCD 300 detector). Images were processed and scaled with HKL2000.[28] Data were collected to a resolution of 1.0 Å (Table 4).

Diffraction Data Collection for the Quasi-Racemate

For low temperature data collection, selected crystals of the L-Se-sfAFP/D-sfAFP quasi-racemate were transferred to the cryoprotectant (reservoir solution plus 20% (v/v) glycerol) for a few seconds and flash-frozen in liquid nitrogen. Three-wavelengh anomalous dispersion X-ray data were collected at 100K at the Advanced Photon Source, Argonne National Laboratory, beamline 23ID equipped with a MARCCD 300 detector). Images were processed and scaled with HKL2000.[28] Data were collected to a resolution of 1.2 Å (Table 1).

Determination of the sfAFP X-ray Structure from the Quasi-Racemate

The L-Se-sfAFP/D-sfAFP quasi-racemate structure was solved by MAD phasing.[23] The heavy atom search was performed in SHELXD[29] using X-ray data collected at 12.6607 keV peak energy and 3.5 Å resolution cutoff. MAD phasing and density modification were performed with use of SOLVE and RESOLVE, respectively.[30] The model building performed with the RESOLVE basic script was complete with use of ARP/WARP program.[31] The electron density and model examinations were done using TURBO-FRODO.[32] The restrained positional and anisotropic B-factor refinement was done in REFMAC5.[33] R_{free} was monitored by setting aside 5% of the reflections as a test set. Crystal data and X-ray data statistics for the quasi-racemate data set are listed in Table 1. Backbone rmsd (Å) deviations between the enantiomeric protein molecules in the quasi-racemate are presented in Table 2. The coordinates and structure factors have been deposited in the Protein Data Bank with entry code 3BOG.

Determination of the X-ray Structure of L-sfAFP

The structure was determined by molecular replacement using the program MOLREP[34] with the L-Se-sfAFP molecule from the quasiracemate as a starting model. The rigid body refinement, restrained positional and anisotropic *B*-factor refinement, was done in REFMAC5.[33] The electron density and model examinations were done using TURBO-FRODO.[32] R_{free} was monitored by setting aside 5% of the reflection as a test set. Crystal data and X-ray data statistics for the L-sfAFP data set are listed in Table 3. The coordinates and structure factors have been deposited in the Protein Data Bank with entry code 2PNE.

Determination of the X-ray Structure of the {L-sfAFP+D-sfAFP} True Racemate

Since the crystal packing of the quasi-racemate is the same as the true racemate, we proceeded directly to rigid-body refinement with two L-fAFP monomers in the asymmetric unit (chains A and B) oriented as for the quasi-racemate. The rigid body refinement, restrained positional and anisotropic *B*-factor refinement, was done in REFMAC5.[33] The electron density and model examinations were done using TURBO-FRODO.[32] R_{free} was monitored by setting aside 5% of the reflection as a test set. Crystal data and X-ray data statistics for the racemate data set are listed in Table 4. The coordinates and structure factors have been deposited in the Protein Data Bank with entry code 3BOI.

Acknowledgment

We thank W. DeGrado (U. Penn) for suggesting that we work on the snow flea antifreeze protein. We thank T. Sosnick, V. Pingali, V. Torbeev, Li-Wei Hung, C. Weeks, S. Anderson, N. Sanishvili, M. Becker, and A. Moglich for useful discussions and technical advice. Z.G. was supported by the University of Chicago NIH Training Program in Physical and Chemical Biology. This research was supported by the Office of Science (BER), U.S. Department of Energy, Grant No. DEFG02-07ER64501 to S.B.H.K., and by the National Institutes of Health, Grant No. R01 GM075993 to SBHK. The use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. DE-AC02-06CH11357. Portions of this work were performed at the SBC-CAT and GM/CA CAT located at Sectors 19BM and 23ID, respectively, of the Advanced Photon Source. SBC-CAT is supported by the U.S. Department of Energy, Office of Biological and Environmental Research, under Contract No. W-31-109-ENG-38. Use of the GM/CA CAT Sector 23-ID was funded with Federal funds from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Science (Y1-GM-1104).

References

1. Davies PL, Baardsnes J, Kuiper MJ, Walker VK. Philos. Trans. R. Soc. London, Ser. B 2002;357:927–935. [PubMed: 12171656]
2. DeVries AL, Wohlschlag DE. Science 1969;163:1073–1075. [PubMed: 5764871]
3. DeVries AL, Komatsu SK, Feeney RE. J. Biol. Chem 1970;245:2901–2908. [PubMed: 5488456]
4. DeVries AL. Science 1971;172:1152–1155. [PubMed: 5574522]
5. Prabhu N, Sharp K. Chem. Rev 2006;106:1616–1623. [PubMed: 16683747]
6. Graham LA, Davies PL. Science 2005;310:461. [PubMed: 16239469]
7. Lin FH, Graham LA, Campbell RL, Davies PL. Biophys. J 2007;92:1717–1723. [PubMed: 17158562]
8. Schnolzer M, Kent SBH. Science 1992;256:221–225. [PubMed: 1566069]
9. Dawson PE, Muir TW, Clark-Lewis I, Kent SBH. Science 1994;266:776–779. [PubMed: 7973629]
10. Dawson PE, Kent SB. Annu. Rev. Biochem 2000;69:923–960. [PubMed: 10966479]
11. Bang D, Kent SB. Angew. Chem., Int. Ed. Engl 2004;43:2534–2538. [PubMed: 15127445]
12. Johnson EC, Kent SB. J. Am. Chem. Soc 2006;128:6640–6646. [PubMed: 16704265]
13. Durek T, Torbeev VY, Kent SB. Proc. Natl. Acad. Sci. U.S.A 2007;104:4846. [PubMed: 17360367]
14. Torbeev VY, Kent SB. Angew. Chem., Int. Ed 2007;46(10):1667–1670.
15. Pentelute BL, Gates ZP, Dashnau JL, Vanderkooi JM, Kent SBH. J. Am. Chem. Soc 2008;130:9702–9707. [PubMed: 18598026]
16. Wukovitz SW, Yeates TO. Nat. Struct. Biol 1995;2(12):1062–1067. [PubMed: 8846217]
17. Pellegrini M, Wukovitz SW, Yeates TO. Proteins 1997;28:515. [PubMed: 9261867]
18. Zawadzke LE, Berg JM. Proteins 1993;16:301–305. [PubMed: 8346193]
19. Hung LW, Kohmura M, Ariyoshi Y, Kim SH. J. Mol. Biol 1999;285:311–321. [PubMed: 9878408]
20. Toniolo C, Peggion C, Crisma M, Formaggio F, Shui X, Eggleston DS. Nat. Struct. Biol 1994;1:908–914. [PubMed: 7773780]
21. Patterson WR, Anderson DH, DeGrado WF, Cascio D, Eisenberg D. Protein Sci 1999;8:1410. [PubMed: 10422829]
22. Berg JM, Goffeney NW. Methods Enzymol 1997;276:619–627. [PubMed: 9048383]
23. Hendrickson WA, Ogata CM. Macromol. Crystallogr., Pt A 1997;276:494–523.
24. Stout, GH.; Jensen, LH. X-Ray Structure Determination: A Practical Guide. Vol. 2nd ed.. New York: Wiley; 1989.
25. Liou YC, Tocilj A, Davies PL, Jia Z. Nature 2000;406:322–324. [PubMed: 10917536]
26. Leinala EK, Davies PL, Jia Z. Structure 2002;10:619–627. [PubMed: 12015145]
27. Schnolzer M, Alewood P, Jones A, Alewood D, Kent SBH. Int. J. Peptide Res. Ther 2007;13:31–44.
28. Otwinowski Z, Minor W. Mocromol. Crystallogr., Pt A 1997;276:307–326.
29. Schneider TR, Sheldrick GM. Acta Crystallogr., Sect. D 2002;58:1772–1779. [PubMed: 12351820]
30. Terwilliger TC, Berendzen J. Acta Crystallogr., Sect. D 1999;55:849–861. [PubMed: 10089316]
31. Perrakis A, Harkiolaki M, Wilson KS, Lamzin VS. Acta Crystallogr, Sec. D 2001;57:1445–1450.

32. Cambillau, C.; Roussel, A. Turbo Frodo. Marseille, France: Université Aix-Marseille II; 1997. Version OpenGL.1

33. Murshudov GN, Vagin AA, Dodson EJ. Acta Crystallogr, Sect. D 1997;53:240–255. [PubMed: 15299926]

34. Vagin A, Teplyakov A. J. Appl. Crystallogr 1997;30:1022–1025.

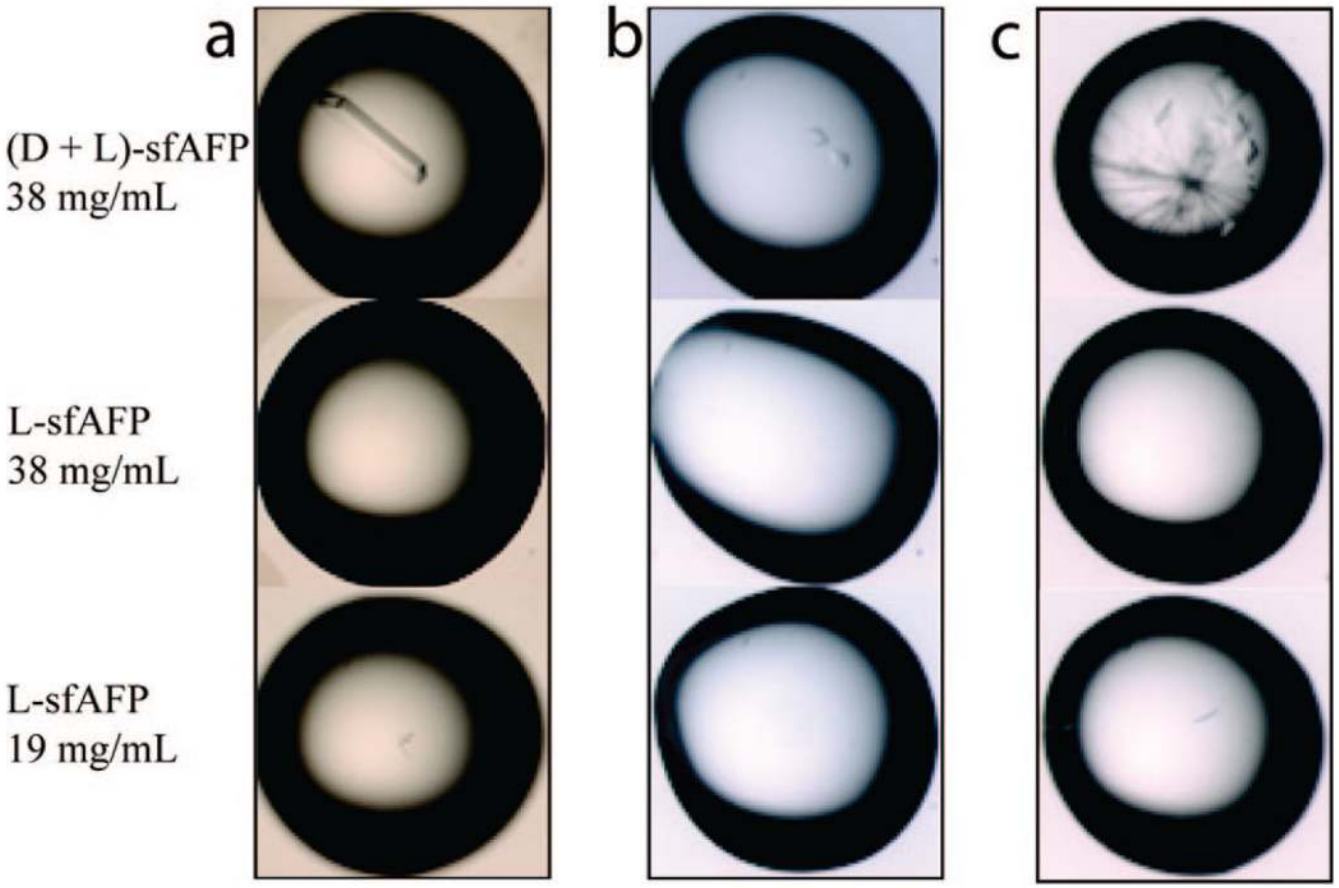

Figure 1.

A direct comparison of the results for attempted crystallization of the racemic protein mixture [D-sfAFP + L-sfAFP] or the single enantiomer L-sfAFP under three sets of conditions. Protein crystals appeared rapidly from a solution containing a racemic mixture of D-sfAFP and L-sfAFP, but not for L-sfAFP alone, when screened under a standard set of crystallization conditions. Representative examples are shown in this Figure. X-ray diffraction was used to verify that the crystals were in fact protein. The crystallization conditions shown are a subset of the main findings reported here and are as follows: (a) 0.1 M bicine pH = 9.0, 2% v/v 1,4-dioxane, 10% w/v polyethylene glycol 20,000; (b) 0.1 M bis-Tris pH = 6.5, 1.15 M ammonium sulfate; (c) 0.1 M Tris pH = 8.5, 25% w/v polyethylene glycol 3350. All crystallization trials were carried out at room temperature (~23 °C).

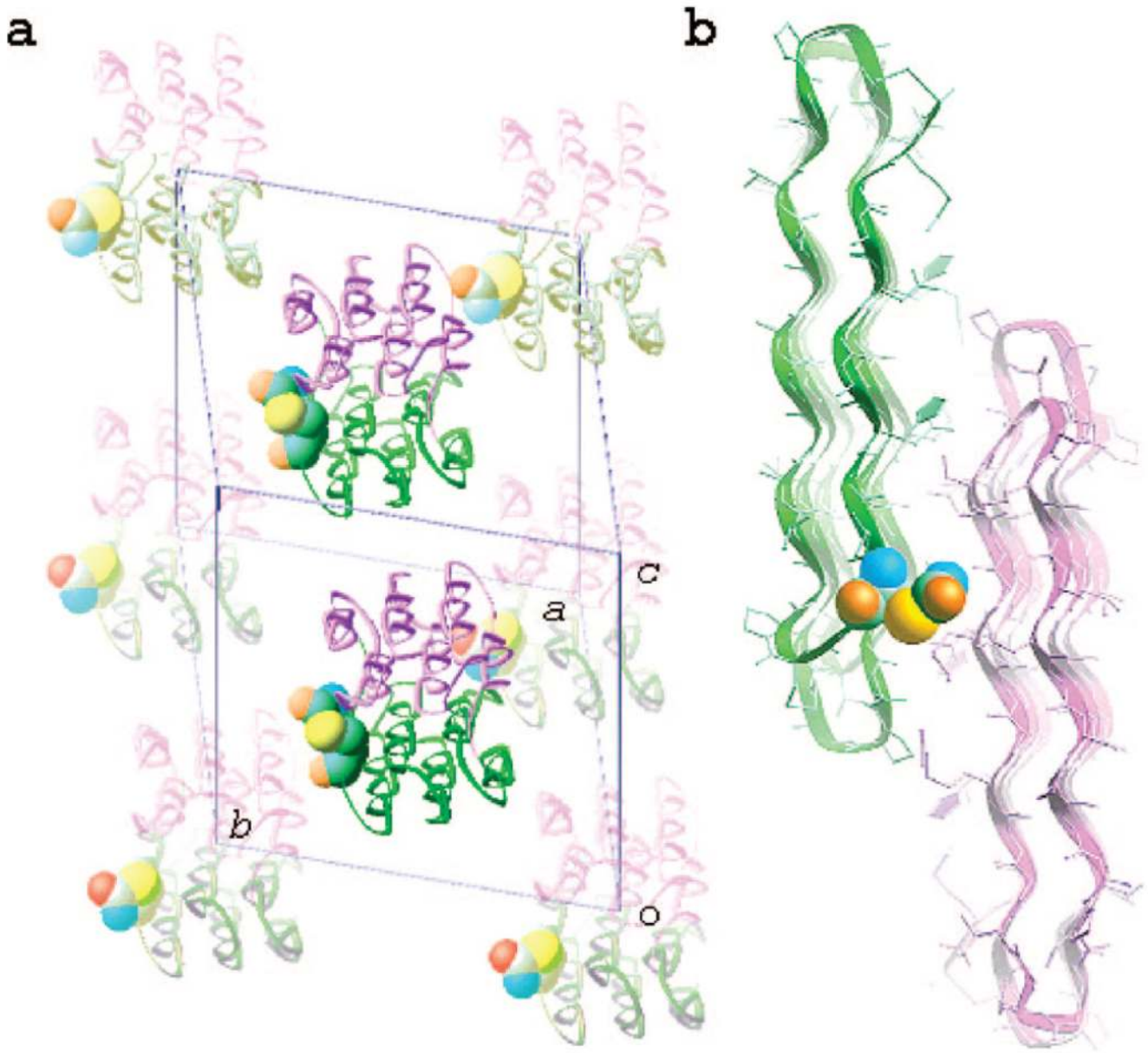

Figure 2.

Quasi-racemate crystal packing in space group *P*1, showing the D-sfAFP molecule (purple ribbon) and the L-Se-sfAFP (green ribbon) with the "pseudo-Se-Gln" at position 11 shown as CPK solid spheres colored by atom type (C, green; O, red; N, cyan; Se, yellow). (a) Unit cell with two D-sf AFP/L-Se-sfAFP pairs located at positions 0, 1/2b, 1/2c, and 0,0,0, in the asymmetric unit. The D-sfAFP contains right-handed polyproline type II (PPII) helices; the L-Se-sfAFP contains left-handed PPII helices. (b) View of one D-sfAFP/L-Se-sfAFP pair located at 0, 1/2b, 1/2c. Side chains are included as stick representations.

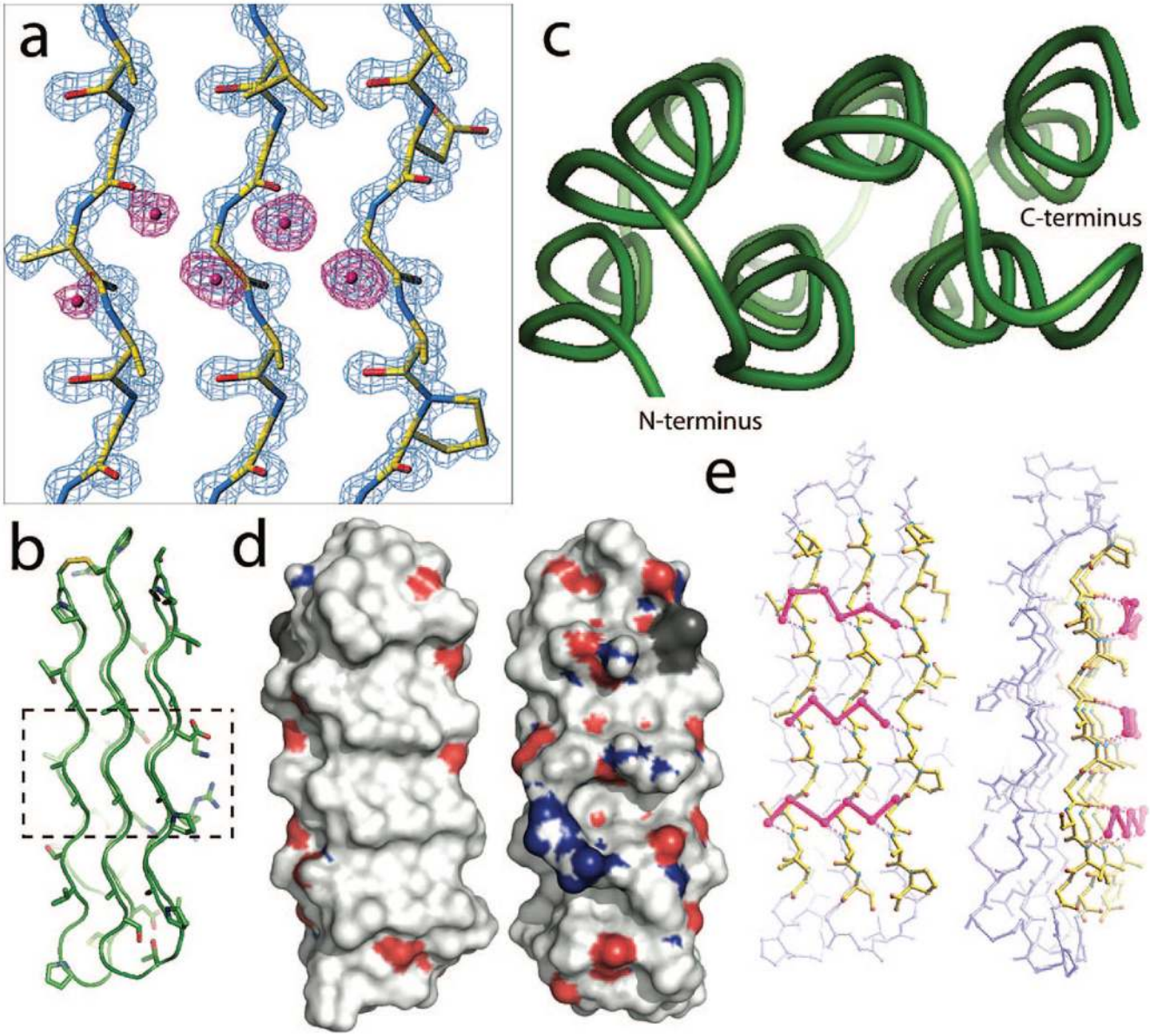

Figure 3.

The X-ray structure of snow flea antifreeze protein (sfAFP). Atomic resolution (1.2 A) X-ray structure of sfAFP, obtained from quasi-racemate crystals of D-sfAFP and L-Se-sf AFP in the *P*1 space group. (a) SigmaA-weighted $2F_o - F_c$ electron density map (blue, 1.5 sigma level) of the proposed ice-binding surface encompassing residues 17–21, 47–51, 74–78. Five ordered water molecules are shown (magenta spheres) with electron density shown at 1σ. (b) Cartoon of the backbone fold of sfAFP. Amino acid side chains are shown as sticks. The dashed box indicates the region shown in (a) above. (c) The main structural elements of sfAFP are six antiparallel polyproline type II (PPII) (left-handed) helices, stacked in two groups and joined by five reverse turns and interlocked by a complex hydrogen bond network. The N-terminal half of the protein contains two intramolecular disulfide bonds between residues Cys1-Cys28 and Cys13-Cys43. (d) Surface map highlighting sfAFP's amphiphilic character. A 180° rotation (along the long axis) shows the contrasting surfaces: hydrophobic/apolar (left); hydrophilic/polar (right). (e) Two views related by a 90° rotation of highly ordered first shell

water molecules (shown in magenta) interacting with the sfAFP hydrophobic face backbone (dashed magneta lines). Water molecules (magenta spheres) are within hydrogen-bonding distance (2.7–3.0 Å) of the sfAFP backbone amide carbonyls and nitrogens, and of one another (solid magnenta lines). The spacing between water molecules is 4.5 Å and 9.2 Å along the short and long sfAFP axes, respectively. X-ray statistics are given in Table 1. Backbone rmsd deviations between molecules of L-sfAFP and D-sfAFP are given in Table 2.

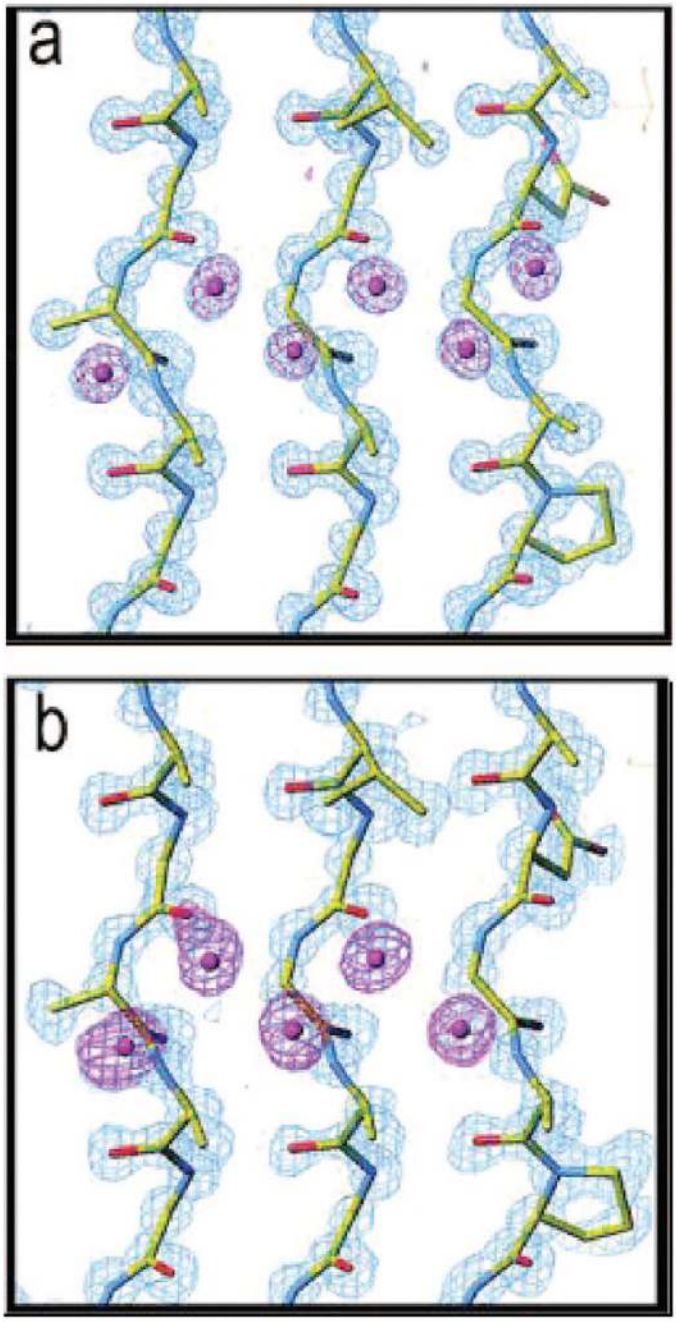

Figure 4.

SigmaA-weighted $2F_o - F_c$ electron density map (blue, 1.5σ level) of the proposed ice-binding surface encompassing residues 17–21, 47–51, 74–78 for: (a) $P2_1$ L-sfAFP; (b) $P\bar{1}$ {D-sfAFP + L-sfAFP}. Ordered water molecules with electron density at 1σ are shown (magenta spheres). Five ordered water molecules are conserved in all structures. An additional water molecule is observed in the $P2_1$ structure; this water is absent in the racemic crystals because of packing interactions.

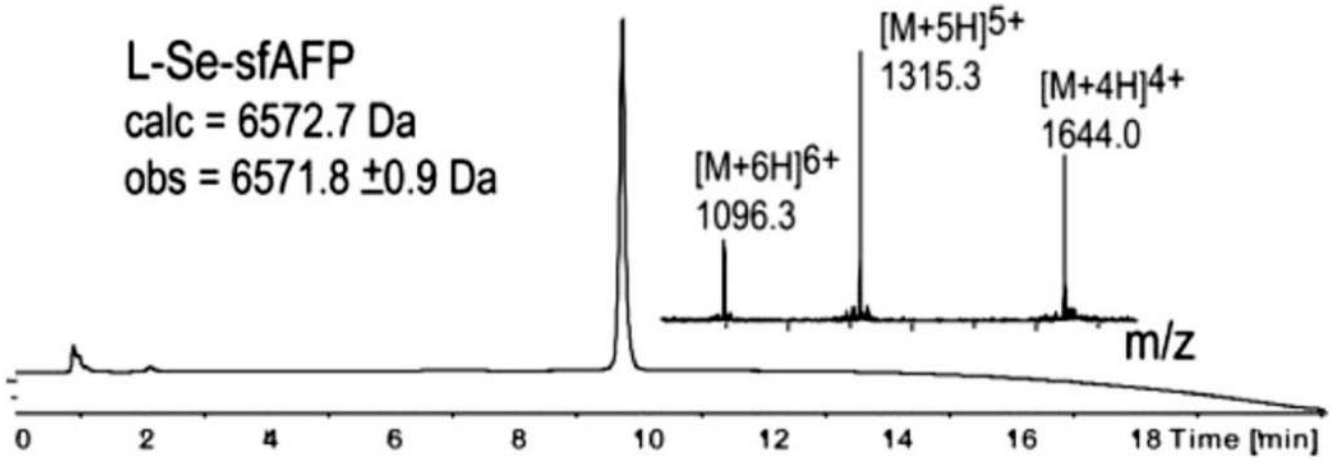

Figure 5.

Analytical LC-MS data for synthetic folded L-Se-sfAFP. The chromatographic separations were carried out on a Vydac C4 2.1 mm × 150 mm column using a linear gradient of 1–61% buffer B over 15 min (buffer *A* = 0.1% TFA in H_2O; buffer *B* = 0.08% TFA in acetonitrile). The inset is the ESMS summed over the entire LC peak. Calculated masses were based on average isotope composition.

Cys[1]-Lys-Gly-Ala-Asp-Gly-Ala-His-Gly-Val[10]-Asn-Gly-Cys-Pro-Gly-Thr-Ala-Gly-Ala-Ala[20]-
Gly[21]-Ser-Val-Gly-Gly-Pro-Gly-Cys-Asp-Gly[30]-Gly-His-Gly-Gly-Asn-Gly-Gly-Asn-Gly-Asn[40]-
Pro[41]-Gly-Cys-Ala-Gly-Gly-Val-Gly-Gly-Ala[50]-Gly-Gly-Ala-Ser-Gly-Gly-Thr-Gly-Val-Gly[60]-
Gly[61]-Arg-Gly-Gly-Lys-Gly-Gly-Ser-Gly-Thr[70]-Pro-Lys-Gly-Ala-Asp-Gly-Ala-Pro-Gly-Ala[80]-
Pro[81]

Scheme 1.
Predicted 81-Residue Glycine-Rich Amino Acid Sequence of sfAFP [6]

Table 1

X-ray Data Collection and Refinement Statistics for the Quasi-Racemate [D-sfAFP+L-Se-sfAFP]

	Crystal 1		
Space group	P1		
Cell parameters	a = 28.64 b = 32.36 c = 59.70 α = 88.62° β = 89.31° γ = 72.84°		
	Peak	*Inflection*	*Remote*
Data Collection Statistics:			
Beamline/Detector	APS 23ID-D/MARCCD	APS 23ID-D/MARCCD	APS 23ID-D/MARCCD
Wavelength (Å)	0.9793	0.9794	0.9493
Resolution range (Å)	50.00 − 1.20(1.24 − 1.20)[a]	50.00 − 1.20 (1.24 − 1.20)	50.00 − 1.20 (1.24 − 1.20)
Reflections measured	856855	426808	422666
Unique reflections measured[b]	60583	60655	60754
Completeness (%)	94.7(91.7)	94.7(91.4)	95.0 (92.0)
Average Redundancy	8.0 (7.9)	4.0 (3.9)	4.0 (4.0)
<I>/<σ(I)>	37.9 (9.9)	33.8 (6.7)	39.6(11.4)
R_{merge}	0.068(0.21)	0.048 (0.18)	0.041(0.11)
Refinement:			
Resolution range (Å)	27.37 − 1.20(1.23 − 1.20)		
Reflections used (working set)	57496 (4052)		
Reflections used (test set)	3064 (208)		
Completeness (working + test set, %)	94.7 (90.5)		
R_{work}	0.160(0.137)		
R_{free}	0.192(0.195)		
Rms deviations			
Bond lengths (Å)	0.016		
Bond angles (°)	1.6		
Average B factor (overall, $Å^2$)	7.8		
Ramachandran Plot Statistics:			
Most favored (%)	93.2 (L), 91.9(D)		
Additionally allowed (%)	6.8 (L), 8.1 (D)		
Generally allowed (%)	0 (L), 0 (D)		

[a] Values for highest resolution shell are given in parentheses.

[b] Friedel mates merged.

Table 2

Backbone rmsd (Å) Deviations between Molecules in the Quasi-Racemate: Molecules A and B are L-Se-sfAFP, and Molecules C and D Are Inverted D-sfAFP

	A	B	C-inverted	D-inverted
A		0.520	0.165	0.514
B			0.564	0.203
C-inverted				0.540

Table 3

X-ray Data Collection and Refinement Statistics for L-sfAFP

Space group	$P2_1$
Cell parameters	a = 16.70 Å
	b = 74.28 Å
	c = 17.69 Å
	β = 102.2°
Data Collection Statistics:	
Beamline/Detector	APS 19BM/SBC2
Wavelength (Å)	0.9792
Resolution range (Å)	50.00 − 0.95 (0.98 − 0.95)[a]
Reflections measured	241662
Unique reflections measured	23972
Completeness (%)	90.6 (46.0)
Average redundancy	3.3 (1.3)
<I>/<σ(I)>	17.9 (3.0)
R_{merge}	0.059 (0.25)
Refinement Statistics:	
Resolution range (Å)	17.29 − 0.98(1.01 − 0.98)
Reflections used (working set)	21784
Reflections used (test set)	1173
Completeness (working + test set, %)	95.5 (83.6)
R_{work}	0.140 (0.303)
R_{free}	0.164 (0.255)
Rms deviations	
Bonds (Å)	0.017
Angles (degrees)	1.7
Average B factor (overall, $Å^2$)	6.6
Ramachandran Plot Statistics:	
Most favored (%)	94.6
Additionally allowed (%)	5.4
Generally allowed (%)	0

[a]Values for highest resolution shell are given in parentheses.

Table 4

X-ray Data Collection and Refinement Statistics for *Racemic* sfAFP

	D, L-sfAFP
Space group	P1<bar>
Cell parameters	a = 28.60 Å
	b = 32.40 Å
	c = 59.85 Å
	α = 88.69°
	β = 89.19°
	γ = 73.41°
Data Collection Statistics:	
Beamline/Detector	APS 23ID-B/MARCCD300
Wavelength (Å)	0.9792
Resolution range (Å)	50.00 − 0.95 (0.98 − .95)[a]
Reflections measured	974696
Unique reflections measured	101192
Completeness (%)	78.2(18.9)
Average redundancy	4.3 (3.3)
<I>/<σ(I)>	17.4(3.6)
R_{merge}	0.084 (0.42)
Refinement Statistics:	
Resolution range (Å)	25.00 − 1.00(1.03 − 1.00)
Reflections used (working set)	92050
Reflections used (test set)	4909
Completeness (working + test set, %)	87.1 (43.7)
R_{work}	0.218(0.229)
R_{free}	0.233 (0.297)
Rms deviations	
Bonds (Å)	0.014
Angles (degrees)	1.6
Average B factor (overall, $Å^2$)	7.9
Ramachandran Plot Statistics:	
Most favored (%)	93.2
Additionally allowed (%)	6.8
Generally allowed (%)	0

[a] Values for highest resolution shell are given in parentheses.