Perplexing cooperative folding and stability of a low-sequence complexity, polyproline 2 protein lacking a hydrophobic core

Zachary P. Gates[a,1], Michael C. Baxa[b,c,1], Wookyung Yu[b,2], Joshua A. Riback[c,d], Hui Li[b], Benoît Roux[a,b,c], Stephen B. H. Kent[a,b,c,3], and Tobin R. Sosnick[b,c,3]

[a]Department of Chemistry, The University of Chicago, Chicago, IL 60637; [b]Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637; [c]Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL 60637; and [d]Graduate Program in Biophysical Sciences, The University of Chicago, Chicago, IL 60637

The burial of hydrophobic side chains in a protein core generally is thought to be the major ingredient for stable, cooperative folding. Here, we show that, for the snow flea antifreeze protein (sfAFP), stability and cooperativity can occur without a hydrophobic core, and without α-helices or β-sheets. sfAFP has low sequence complexity with 46% glycine and an interior filled only with backbone H-bonds between six polyproline 2 (PP2) helices. However, the protein folds in a kinetically two-state manner and is moderately stable at room temperature. We believe that a major part of the stability arises from the unusual match between residue-level PP2 dihedral angle bias in the unfolded state and PP2 helical structure in the native state. Additional stabilizing factors that compensate for the dearth of hydrophobic burial include shorter and stronger H-bonds, and increased entropy in the folded state. These results extend our understanding of the origins of cooperativity and stability in protein folding, including the balance between solvent and polypeptide chain entropies.

protein folding | cooperativity | kinetics | PP2 | hydrogen bonding

The basis of protein-folding stability and cooperativity remains a topic of great interest (1, 2). Most studies have focused on proteins with hydrophobic cores containing α-helices and β-sheets. Here, we study the folding of snow flea antifreeze protein (sfAFP), a globular protein that lacks these features. The 81-residue protein has a novel fold that is distinct from other proteins (3–6), containing only polyproline 2 (PP2) helices and turns with a core filled with H-bonds and no hydrophobic groups (Fig. 1).

The H-bond network between the PP2 helices (6), $PP2_{1–6}$, requires close packing, which would be precluded by the presence of a C_β atom. Thus, the PP2 helices are defined by glycine (Gly) repeats (GXX or GGX, where X is any other amino acid), which enable the hydrogen-bond network. As a consequence, the sfAFP molecule has a low-complexity glycine-rich sequence [37/81 Gly residues, or 46%; typical is 6, or 7% (7)]. The sfAFP core is well packed with essentially no interior void volumes, even when probed using a 0.5-Å radius sphere (8). sfAFP also contains two disulfide bonds, C1–C28 and C13–C43.

Notably, no side chains are buried in sfAFP's core (2). All 12 hydrophobic side chains (V_4, K_2, and P_6) are surface exposed. Upon folding, sfAFP buries about 20 $Å^2 \cdot res^{-1}$ of hydrophobic surface area compared with an average of 50 $Å^2 \cdot res^{-1}$ for a set of 34 proteins of similar size (Fig. 2A). The difference equates to a decrease in stability of ~1 or 1.4 $kcal \cdot mol^{-1} \cdot res^{-1}$, assuming a surface tension coefficient of $\gamma \sim 34$ or 47 $cal \cdot mol^{-1} \cdot Å^{-2}$ based on classical (9) or Flory–Huggins theory (10), respectively. Even the lower bound of 1 $kcal \cdot mol^{-1} \cdot res^{-1}$ represents a significant stability loss, and it is not obvious which energetic terms could compensate for the reduction in hydrophobic burial.

Because hydrophobic burial is generally regarded as the basis of protein stability and folding cooperativity (11), we investigated sfAFP's folding behavior. We characterize the stability and folding kinetics using circular dichroism (CD), fluorescence, and small-angle

X-ray scattering (SAXS). We find that sfAFP folds in a manner consistent with cooperative two-state folding with an energy surface dominated by the native and unfolded states. We propose that an unusual match between the dihedral angle propensities in the unfolded state ensemble and the secondary structure of the native state, increased native state backbone entropy, and stronger native state H-bonds are major contributors to the stability of the protein molecule.

Results

Thermal Denaturation of sfAFP. sfAFP was prepared using total chemical synthesis (12) with the two disulfide bonds formed at all times, except for the SAXS measurements. The CD spectrum at 4 °C, pH 8.0, is characteristic of PP2 helices, having a deep minimum at 195 nm and a positive peak at 217 nm (Fig. 3A). The positive peak disappeared upon heating to 60 °C and was restored upon overnight incubation at 4 °C. The unfolded CD spectrum is typical of what is observed in other proteins upon thermal denaturation. Although often termed "random coil," this state represents an ensemble of backbone geometries with the (ϕ,ψ) dihedral angles predominantly in the PP2 basin and lesser populations in the α_R, α_L, and β basins (13–16).

Thermal denaturation produced a sigmoidal curve with a mid-point T_{mid} of 32.5 ± 0.2 °C at a thermal ramp rate of 1 °C·min⁻¹. The thermal transition was independent of protein concentration over a range of 85–300 μM (Fig. 3B and Fig. S1B), indicating that the protein remained monomeric over the concentration range

---

Significance

The basis of protein-folding cooperativity and stability elicits a variety of opinions, as does the existence and importance of possible residual structure in the denatured state. We examine these issues in a protein that is striking in its dearth of hydrophobic burial and its lack of canonical α and β structures, while having a low sequence complexity with 46% glycine. Unexpectedly, the protein's folding behavior is similar to that observed for typical globular proteins. This enigma forces a reexamination of the possible combination of factors that can stabilize a protein.
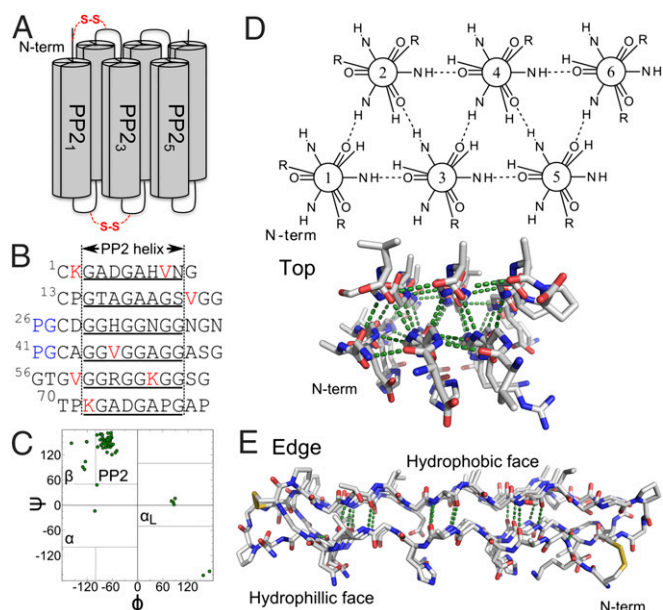

Fig. 1. sfAFP structure (PDB ID code 2PNE). (*A*) Connectivity with the disulfide bonds is noted in red. (*B*) Sequence with helical regions (underlined), valines and lysines (red), and the two PG turns (blue). (*C*) Ramachandran map of native structure with four major basins noted. (*D*) Top and (*E*) edge view highlighting the internal H-bond network and the flat hydrophilic and hydrophobic faces.

studied. Faster thermal ramps brought forth a marked increase in T_{mid}, being 16 °C higher at a ramp rate of 4 °C·min^{-1} compared with 0.25 °C·min^{-1} (Fig. S1*A*). The increased stability is a manifestation of the unfolded state having less time to fully equilibrate due to slow unfolding ($k_u \sim 10^{-5}$ s^{-1} at 4 °C) and *cis/trans* isomerization for the two *cis* and four *trans* Xaa-Pro peptide bonds in the native fold (2).

In the unfolded state, Xaa-Pro peptide bonds isomerize on the minute timescale (17) and equilibrate with a ratio of [*trans*]/[*cis*] $\sim$ 5, depending on the neighbor's identity (18). The net reduction in stability due to this isomerization effect is the sum of individual values, $\Delta G_{cis-trans} = 1.4 + 1.3 + 0.08 + 0.07 + 2 \cdot 0.04 = 2.9$ kcal·mol^{-1} (18). Although the nonequilibrium caused by the prolines compromises the validity of obtaining thermodynamic quantities, we have fit the transition at different temperature ramps and extrapolated the parameters to the limit of an infinitely fast ramp where isomerization in the unfolded state can be ignored (Fig. S1*A*, *Inset*). After extrapolation, we find that ΔH, $T\Delta S$, and ΔG are 38 ± 3, 32 ± 3, and 5.9 ± 0.6 kcal·mol^{-1}, respectively, at $T = 4$ °C (Table S1).

Monitoring Denaturation with Multiple Probes. To permit the monitoring of folding using both CD and fluorescence, two sfAFP analogs were prepared, sfAFP^{H32W} and sfAFP^{R62W}. Thermal denaturation measurements indicated that the analogs had similar stability to wild-type sfAFP ($T_{mid} = 27.6 \pm 0.1$ and 25.9 ± 0.1 °C for sfAFP^{H32W} and sfAFP^{R62W}, respectively, versus $T_{mid} = 32.5 \pm 0.2$ °C for WT sfAFP at a thermal ramp rate of 1 °C·min^{-1}) (Fig. S1 *C and D*). For the Trp variants, the CD$_{219nm}$ and fluorescence signals ($\lambda_{excite} = 285$ nm) changed concurrently with temperature and yielded the same T_{mid} values. This equivalence is consistent with a cooperative folding transition, as fluorescence is a local probe of the Trp environment, whereas CD$_{219nm}$ reports on the global formation of the PP2 helices. This finding provides initial evidence that the free-energy surface is well approximated near the folding midpoint as a double well potential with an unfolded state and a folded state.

Folding Kinetics and Cooperativity. Kinetic measurements were conducted on sfAFP^{R62W} to determine whether the energy landscape was approximated as a double well potential under strongly folding and unfolding conditions. The CD and fluorescence signals were simultaneously acquired at 4 °C in a CD spectrometer using the same wavelength ($\lambda = 226$ nm; Fig. 4*A* and Fig. S2) and guanidinium chloride (GdmCl) as a denaturant. Above 4 M GdmCl, unfolding rates were measured using a stopped-flow apparatus monitoring only fluorescence. Measuring folding rates independent of nonnative Pro isomers required "double-jump" measurements (17). During refolding, the CD and fluorescence signals tracked each other across different folding conditions. There was no evidence of a lag phase indicative of a folding intermediate, nor was any measurable subsecond change in CD amplitude observed before the first time point. The lack of a "burst" phase indicates that no measureable amount of secondary structure forms in a kinetic step before the folding to the native state.

The CD and fluorescence traces were fit to a single-exponential function having nearly identical rates (Fig. S2 and Table S2). This agreement indicates that the two signals are tracking processes that are limited by the same kinetic barrier and that folding occurs in a two-state manner without the accumulation of intermediates. The fit to the data improved slightly when allowing for a second phase for both probes of up to 10 or 20% of the total signal for unfolding or folding, respectively (Fig. S2 and Table S3). A second phase observed by CD and fluorescence, both in rate and amplitude, suggests heterogeneity in the starting state rather than the accumulation of an intermediate. The slow minor folding phase could be due to *cis–trans* Pro isomerization occurring during the minute-long refolding process, whereas the fast minor unfolding phase may be due to a nonnative proline isomer in an otherwise-folded structure.

Below 4.5 M GdmCl, the major folding and unfolding phases could be fit with a two-state U↔N model, with the activation free energies for folding and unfolding being linear with denaturant concentration (19). However, the unfolding region of the associated chevron plot exhibits a change in slope at ∼4.5 M GdmCl (Fig. 4*B*). Such kinks have been observed in the unfolding of other apparent two-state proteins, and analyzed using a reaction scheme (Fig. 4*C*) with two sequential transition states, TS1 and TS2, and an intervening high-energy intermediate state (I*) of
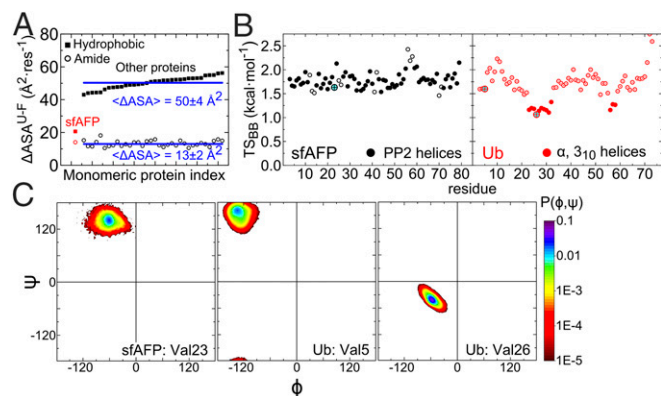

Fig. 2. Properties of sfAFP and other proteins. (*A*) Hydrophobic and amide surface burial for sfAFP and similar-sized proteins. The calculation of sfAFP's hydrophobic (■) and amide burial (○) assumes an unfolded state containing the two disulfide bonds (values increase by 29–30% when the disulfide bonds are removed). (*B*) Native-state backbone entropy for sfAFP and ubiquitin (Ub). Larger circles denote residues highlighted in *C*. (*C*) Ramachandran probability distributions from MD simulations for a valine in a PP2 helix (sfAFP) and in α and β structures (Ub).
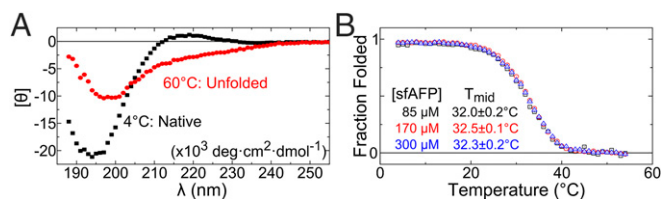
Fig. 3. Thermal denaturation of sfAFP. (*A*) CD spectra of folded (4 °C, black squares) and unfolded sfAFP^H32W (60 °C, red circles). (*B*) Normalized thermal denaturation as a function of protein concentration for WT sfAFP.

indeterminable stability (20, 21). Below 4.5 M GdmCl, TS1 is rate-limiting for folding and unfolding. The kink arises from the differential surface burial (denaturant dependence) of TS1 and TS2 (Table S4), which results in TS2 becoming rate limiting for unfolding above 4.5 M GdmCl. Although postulating an I* species complicates the energy surface, the state does not accumulate, and, therefore, folding still appears kinetically two-state.

This kinetic analysis produced a value for the equilibrium stability of sfAFP^R62W, $\Delta G_{eq} = 6.4 \pm 0.6$ kcal·mol^{-1}. Because the folding rates were obtained using a double-jump protocol that maintained the native Pro isomers, the chevron-derived stability reflects the difference between the native state and a denatured state having the native isomers. The chevron analysis also produced an equilibrium *m* value, $m^o = 2.5 \pm 0.2$ kcal·mol^{-1}·M^{-1}, consistent with other proteins. The m^o value is a measure of the surface buried going from the unfolded to the native state (22). sfAFP is predicted to have an $m^o = 1.8$ kcal·mol^{-1}·M^{-1} (23), assuming an unstructured denatured state model (13), which is less than our experimental value. If a rapidly formed intermediate did exist and had escaped our detection, the net m^o would be reduced. Hence, the sizable m^o value obtained from the chevron provides further support that folding is dominated by two energy wells.

SAXS Studies. Poly(Gly) sequences have been observed to form collapsed or aggregated structures (24–28), and so we sought to determine whether a denatured state of sfAFP with 46% Gly content would behave similarly. To this end, we measured the properties of sfAFP^H32W with reduced disulfide bonds using SAXS coupled with inline size exclusion chromatography (SEC) (Fig. 5). Without the two disulfide bonds, the species is not the true unfolded state of sfAFP. However, the reduced species provides an analog accessible under aqueous conditions to ascertain whether sfAFP's Gly content is sufficient to promote collapse or aggregation. The reduced protein eluted from the size exclusion column as a single peak at an elution volume earlier than expected for a globular (i.e., collapsed) protein. Furthermore, the scattering profile was consistent with the Debye model for a random walk (29), indicating that the reduced analog was unfolded with at least one broken disulfide bond (Fig. 5).

The fit of the data with the Debye model over the range $q \cdot R_g \leq 2$ yields a radius of gyration, $R_g = 23.1 \pm 0.2$ Å. The crystal structure R_g is 13.3 Å, whereas the expected R_g for a chemically denatured protein of this size is 25.4 Å. The latter is calculated from the known correlation between chain length (N = 81) and R_g for a self-avoiding random walk (SARW), $R_g = 2N^{0.59}$ (30), when corrected by 5% for the difference between the scaling for an Ala- versus Gly-containing random walk (31). Although this disordered state (under aqueous conditions) is not as expanded as a completely self-avoiding random walk, the SAXS data indicates that the reduced analog is not collapsed, consistent with findings on other unfolded proteins (32–35). These results indicate that high glycine content does not necessarily produce a collapsed globule or aggregation, unlike that observed both experimentally and in simulations using poly(Gly) sequences (24–28). Possibly the

difference lies in the C$_\beta$-containing residues in the heterogeneous sfAFP sequence that serve to restrict an otherwise-favorable collapse of a polypeptide with a high Gly content.

Conformational Entropy. Explicit solvent molecular dynamics (MD) simulations were analyzed to calculate the conformational entropy of sfAFP in its native state for comparison with the model protein ubiquitin (Ub). Entropy is calculated according to $S = -R \Sigma_i P_i \ln P_i$, where P_i is the population in 10° × 10° pixels in the Ramachandran map (36, 37). We find that residues in sfAFP's PP2 helices and in Ub's β-sheets exhibit comparable backbone entropy (Fig. 2B). Helical residues in Ub have tighter (ϕ, ψ) distributions and, hence, significantly lower entropies (Fig. 2C). In entropic terms, the less restricted motions in PP2 helices translates to an entropic benefit of $\Delta S = 0.5$ kcal·mol^{-1}·res^{-1}/(300 K), compared with helical residues.

Our prior studies (36, 37) found that the backbone entropy of glycine in the unfolded state exceeds other residues by 0.3 kcal·mol^{-1}·res^{-1}/(300 K). However, other residue types additionally lose side-chain entropy upon folding. For Ub, the average loss of side-chain entropy for the non-Gly, -Ala, and -Pro residues is $\Delta S = 0.3$ kcal·mol^{-1}·res^{-1}/(300 K). This extra entropy loss of side-chain entropy produces an average net entropy loss upon folding similar to that of glycine. Also, the two disulfide bonds in sfAFP (C1–C28, C13–C43) reduce the conformational entropy of the unfolded state through covalent loop closure, by ~5 kcal·mol^{-1}·(300 K)$^{-1}$ (although possibly less as the loops are intertwined) (38).

H-Bonds. sfAFP has an H-bond content and amide burial level that is typical of similar-sized proteins (Fig. 2A). However, sfAFP's H-bonds are on average 0.17 and 0.09 Å shorter than α-helical and β-sheet H-bonds, respectively (Fig. 6A). The reduced distances suggest that the lack of side chains in the core of sfAFP allows the H-bonds to energy minimize and dictate the inter-PP2 helix separation distance. Consistently, the shortest H-bonds in sfAFP are in the middle of the helices, and longer H-bonds are located near the turns where additional constraints are present. We suspect that side-chain packing and the requirement that the main
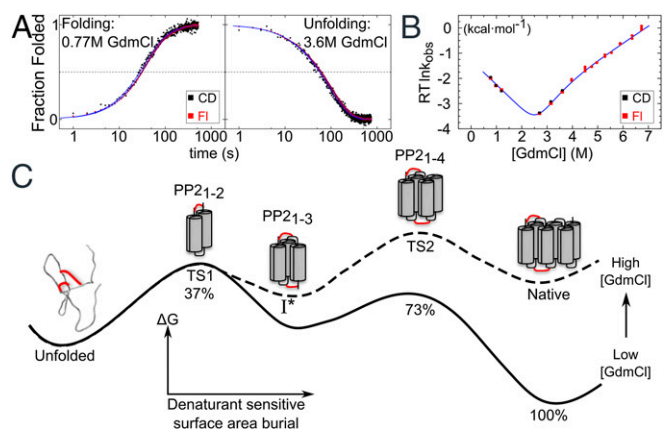

Fig. 4. Folding kinetics of sfAFP. (*A*) Folding of sfAFP^R62W was monitored using CD (black) and Trp fluorescence (red) and fit using a double-exponential model (blue lines). The final GdmCl concentration is noted in each panel. (*B*) Chevron plot fit using a two-state model with two TSs and a high-energy intermediate I*. (*C*) Proposed folding pathway and free-energy surface at low (solid line) and high (dashed line) denaturant concentration where TS1 and TS2 are rate limiting, respectively. The shift in the rate-limiting barrier results in a kink in the chevron plot, which can be explained by the presence of a high-energy I* state. TS1 and TS2 bury 37% and 73% of the denaturant-sensitive surface area, respectively. Only the folded portions of the TS1, I*, and TS2 are shown for simplicity. Red lines indicate disulfide bonds.
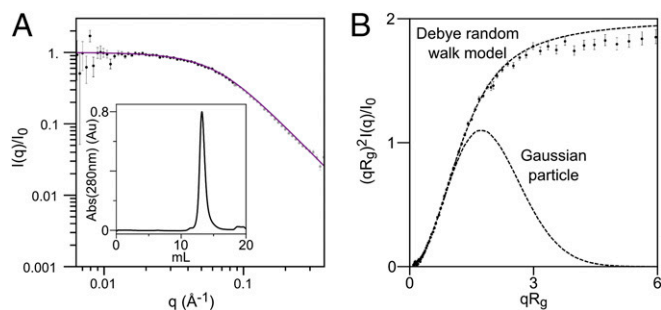
Fig. 5. SAXS on reduced and unfolded sfAFP. (*A*) Data are fit to a Debye random walk model out to $qR_g \leq 2$ (black versus gray data points). (*Inset*) SEC elution profile. (*B*) Data, presented in the form of a dimensionless Kratky plot, are compared with that for a random walk and a Gaussian particle. The protein in 10 mM DTT was injected onto a SEC column equilibrated with 2 mM DTT, pH 7.4.

chain twist result in slightly longer H-bonds in β-sheets and α-helices compared with those in the core of sfAFP.

Translation of the difference in H-bond length into energetic terms is challenging. As a proxy, electrostatic energy calculations between the amide and carbonyl partial charges with molecular force fields were carried out but exhibited little distance dependence (Fig. S3*A*). Furthermore, the H-bonds in sfAFP appear energetically indistinguishable from the other H-bonds according to this simple analysis based on electrostatics.

As an alternative, quantum-mechanical (QM) calculations using B3LYP density functional theory (DFT) (39) were performed on 29 pairs of *N*-methylacetamide (NMA) molecules having geometries matching the 14 parallel and 15 antiparallel PP2 helical H-bond orientations found in sfAFP. An additional 215 NMA pairs using geometries found in our protein test set (100 α-helical, 100 antiparallel, and 15 parallel β orientations) were included for comparison (Fig. 6*A*, Fig. S3*B*, and Table S5). The H-bond energy was defined as the difference between the NMA pair energy and the sum of the individual energies in the same configuration, $E_{HB} = E_{total} - (E_1 + E_2)$. Three solvent dielectric constants were examined, $\varepsilon = 2$, 4, and 10, which likely covers the value inside a protein (40). On average, the antiparallel PP2 H-bond is ~0.3–0.7 kcal·mol^{-1} more stable than those in other secondary structures. The most stable H-bonds tended to have an H–O distance of 1.85–1.90 Å. Increasing ε from 2 to 4 or 10 stabilizes H-bonds by about 0.8 or 1.3 kcal·mol^{-1}·HB^{-1}, respectively, but has little impact on the ΔE_{HB} between the different secondary structures (Fig. S3*B*).

In addition to amide H-bonds, sfAFP has 15 CH–O=C H-bonds, which are also found in other PP2-containing proteins and polypeptides (41, 42). Although the strength of these interactions is unknown, they are likely to be weak (41, 43), especially considering that in sfAFP the carbonyl oxygen already is involved in a backbone H-bond.

Another possible stabilizing factor is the n→π* interaction between adjacent peptide bonds (44). Simulations have suggested that dipole–dipole interactions between peptide bonds, which correspond to the quantum mechanical n→π* interactions, can explain the self-association of polyglycine peptides (25, 45). The distance between oxygen and carbons is shorter in α than PP2 helices (2.9 and 3.2 Å, respectively), suggesting that n→π* interactions are less significant in sfAFP than in typical helical proteins. Because the stability of helical proteins already can be attributed to hydrophobic burial, we infer that n→π* interactions are unlikely to provide significant stabilization for sfAFP.

Significance. Despite lacking canonical secondary structures or a hydrophobic core, sfAFP both folds cooperatively and is stable. These two properties are related but distinct (2, 46); for example,

a small subpopulation can fold cooperatively to a unique structure even under strongly destabilizing conditions if there are two distinct thermodynamic wells. We first discuss sfAFP's stability.

In 1951, Pauling and others (2, 47, 48) stressed the importance of H-bonds. The overall benefit of H-bonding to stability, however, is unclear as the net number of total protein plus water H-bonds typically remains nearly constant upon folding. In 1959, Kauzmann (49) stressed that the hydrophobic effect is the dominant factor in protein stability, in agreement with more recent thinking (50, 51).

Hydrophobic burial provides a large entropic benefit, $\Delta S_{solvent}$, due to the release of water molecules caging the hydrophobic groups in the unfolded state (50, 52). The experimentally observed entropy loss upon folding often is near zero: $\Delta S_{exp} \sim 0 \sim \Delta S_{solvent} + \Delta S_{conf}$, implying that the gain in solvent entropy gain is largely counterbalanced by the loss of conformational entropy, $\Delta S_{solvent} \sim -\Delta S_{conf}$ (50, 52). Indeed, our estimation of $\Delta S_{conf} \sim 1.4$ kcal·mol^{-1}·res^{-1}/(300 K) (36, 37), is similar to the above estimate of hydrophobic stabilization. However, sfAFP buries only 40% of the hydrophobic surface area of a typical protein. This difference equates to a reduction in stability of ~1 or 1.4 kcal·mol^{-1}·res^{-1} (9, 10). Because sfAFP lacks extensive hydrophobic burial, we expect a reduction in $\Delta S_{solvent}$ but no corresponding change in ΔS_{conf}. Hence, the compensating factors that help stabilize sfAFP are expected to have a significant entropic component.

We propose that three factors largely compensate for the lack of hydrophobic burial in sfAFP: strong H-bonds between PP2 helices, heightened native state entropy, and similarity between the residue-level statistical bias toward PP2 dihedral angles in the unfolded state and the actual PP2 structure in the native state. According to our DFT calculations, sfAFP's 15 antiparallel PP2 H-bonds are stronger by ~0.5 kcal·mol^{-1}·H-bond^{-1} compared with other types of H-bonds (Fig. 6*A*). sfAFP's heightened native-state entropy stabilizes the protein by reducing the loss of entropy upon folding, also achievable through the use of specific amino acids (53). Compared with α-helices, the backbone of residues in sfAFP's PP2 helices are slightly more flexible and sample more area in the Ramachandran map during MD simulations (Fig. 2*C*), which equates to an increased native state entropy worth $\Delta S \sim 0.5$ kcal·mol^{-1}·res^{-1}·(300 K)$^{-1}$.

The third major proposed factor in sfAFP's stability is the similarity between the backbone conformations in the folded and unfolded states. The CD spectrum of the denatured chain (Fig. 3*A*)
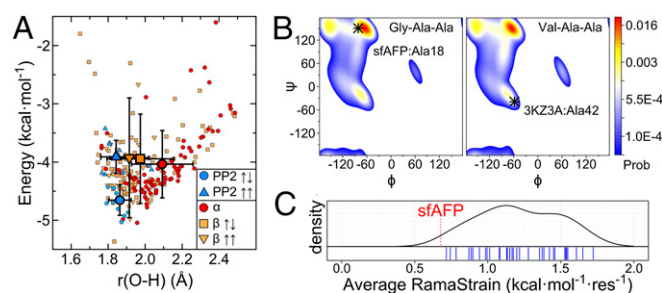

Fig. 6. Proposed major factors that stabilize sfAFP. (*A*) Calculated H-bond energies and distances using DFT for $\varepsilon = 4$ in sfAFP and similar-sized proteins are sorted according to secondary structure. For clarity, two points at (1.60 Å, 0.51 kcal·mol^{-1}) and (1.65 Å, 0.65 kcal·mol^{-1}) lying outside the plotted area are not shown. Larger symbols with error bars denote average values and SDs. (*B*) Ramachandran probability distributions for Ala18 and Ala42 in the unfolded states of sfAFP and λ-repressor, respectively, based on a coil library (59). The native-state conformations are indicated. RamaStrain is the work required to shift the (ϕ, ψ) distribution of the unfolded state to that of the native state. More work is required to adopt a helical conformation compared with a PPII conformation. (*C*) A kernal-density estimate of average RamaStrain calculated for a set of 35 similar sized proteins (individual values denoted with blue tick marks). sfAFP's average RamaStrain is shown with a red dashed line.

indicates that the backbone preferentially adopts (ϕ,ψ) values compatible with PP2 geometry (13, 14, 16, 54–56). The unfolded state's PP2 bias is a statistical property and does not necessitate that the chain have contiguous stretches of PP2 conformers, as observed in the native PP2 helices. Rather, the backbone samples a variety of (ϕ,ψ) angles, and according to the SAXS data, behaves nearly as a SARW when the disulfide bonds are broken.

The PP2 bias is especially strong for chains containing alanine (57) and glycine (24), which are found at high levels in sfAFP's PP2 helices (11 Ala, 33 Gly, out of 68 helical residues). We propose that the strong residue-level bias toward PP2 conformations in the unfolded state reduces the free energy needed to shift its backbone distribution to the native-state distribution. We term this energy "RamaStrain." It has both entropic and enthalpic components. It would be solely entropic if the probability distribution is such that all allowed angles are equally probable (an energy surface with flat-bottom wells). Similar principles have been used to reduce the entropy in the unfolded state through the use of selected substitutions (58).

We evaluated RamaStrain using linkage and an unfolded state (ϕ,ψ) distribution obtained from a structure-based coil library (59), $\Delta G^{\text{RamaStrain}} = -RT \ln P^{\text{unf}}(\phi,\psi)_{\text{nat}}$, where $P^{\text{unf}}(\phi,\psi)_{\text{nat}}$ is the probability of a residue in the unfolded state occupying the native (ϕ,ψ) basin (Fig. 6, Fig. S4, and *SI Methods*). Compared with a test set of 35 comparably sized proteins, sfAFP's RamaStrain is smaller by 0.53 ± 0.28 kcal·mol^{-1}·res^{-1}. This serves as a third factor compensating for sfAFP's lack of hydrophobic burial.

Cooperativity. No generally accepted answer exists to the origin of the all-or-none cooperativity typically observed in the folding of globular proteins. Although hydrophobic burial alone may be insufficient to yield cooperative folding behavior (60, 61), a common presumption is that formation of a hydrophobic core is requisite for two-state folding behavior (11). Simulations and experiments have also stressed the importance of long-range contacts (62, 63). Other factors include multibody energy terms (64), instability of isolated substructures, side-chain packing (11), desolvation (65), and the lack of stable, misfolded structures.

Thermal denaturation of sfAFP produces a sharp unfolding transition with concurrent changes in local (fluorescence) and global (CD) probes, consistent with the native and unfolded states being the only significantly populated species. Similarly, the two probes track each other during the kinetic measurements. Both unfolding and refolding reactions are primarily monophasic, with rate constants that vary linearly with denaturant concentration. The effect of proline isomerization compromises our ability to apply further tests of cooperativity [e.g., comparison of $\Delta H_{\text{van't Hoff}}$ and ΔH_{calor} (66)]. However, the available data provide strong preliminary evidence that sfAFP folds in a single kinetic step, potentially through a high-energy intermediate. Taken together, our findings suggest that a hydrophobic core with specific side-chain packing is not an absolute requirement for cooperative folding behavior.

In sfAFP, the two central helices, PP2$_3$ and PP2$_4$, collectively form H-bonds with all other PP2 helices, with each one individually contacting four helices (Fig. 1D). This level of contact between secondary structures is similar to that found for β-strand and α-helices in similar-sized proteins. The extensive interaction network found in sfAFP is reflected in a residue–residue contact map with many off-diagonal elements and a high relative contact order value of 0.17 (Ub's value is 0.15) (67). These structural intricacies mimic those observed in globular proteins and likely contribute to sfAFP's two-state folding behavior.

We hypothesize the following folding pathway to rationalize the presence of a large free-energy barrier between the native and unfolded states (Fig. 4C). The presence of the two disulfide bonds biases folding to initiate at the N terminus. However, the formation of a single PP2 helix or even a pair of helices with only a single helix–helix interface is likely to be unstable. The β_{Tanford} surface burial

percentages for TS1 and TS2 are $37 \pm 6\%$ and $73 \pm 1\%$, respectively, which suggests that two PP2 helices form in TS1, three helices in the high-energy intermediate I*, and four helices in TS2. The initial unfolding event could be the unfolding of PP2$_6$ as removing this helix is the least disruptive. However, the accompanying gain in entropy for the unfolding of a single PP2 helix comes at the expense of two helix–helix interfaces, making this event net unfavorable. Due to the uphill nature of folding the first two to three helices in TS1 and I*, and the downhill folding of the last two helices from TS2, there is a sizable free-energy barrier on an energy surface dominated by two wells. Further studies are required to evaluate this proposed pathway.

The well-defined native state of sfAFP is a product of its H-bonding pattern, side-chain packing requirements, and turn sequences. The native arrangement of the PP2 helices uniquely positions all C$_\beta$ atoms outward from the core to promote formation of the extensive H-bond network. In addition, the five turns have specific sequence motifs. Four turns contain a Pro, of which two adopt the noncanonical *cis* ω angle. The remaining turn sequence, GGTG, contains two Glys (G56, G58) with (ϕ,ψ) dihedral angles that are strongly disfavored for all other residues. The location of the turns in the sequence also produces nearly equal length PP2 helices and hence maximizes the number of interstrand H-bonds. In sum, the requirement of having Gly at critical positions combined with the strong turn signals helps produce a well-defined native state for sfAFP.

Conclusion

Despite lacking α and β structures and a hydrophobic core, the folding behavior of sfAFP matches those of proteins that have such features. Hence, they are unnecessary for folding cooperativity and stability. sfAFP's two-state behavior can be rationalized by its long range contacts and unique structure due to specific H-bonding requirements and turn motifs. What is less clear are the stabilizing terms that compensate for sfAFP's reduced amount of hydrophobic burial, estimated to provide 1 or 1.4 kcal·mol^{-1}·res^{-1} less stability compared with typical proteins. We propose a major compensation for the lack of burial is that the intrinsic residue-level PP2 bias in the unfolded state reduces the free-energy cost of the backbone adopting the native PP2 geometry by ~ 0.5–1 kcal·mol^{-1}·res^{-1} compared with other, especially helical proteins. In addition, DFT calculations indicate that sfAFP's antiparallel H-bonds are more stable by ~ 0.5 kcal·mol^{-1}·H-bond^{-1} compared with those in other secondary structures. Further studies are required to generate a more complete description of the energetics and folding behavior of this fascinating protein.

Methods

1. Pastore A, Wei G (2015) Editorial overview: Folding and binding: Old concepts, new ideas, novel insights. *Curr Opin Struct Biol* 30:iv–vi.

2. Rose GD, Fleming PJ, Banavar JR, Maritan A (2006) A backbone-based theory of protein folding. *Proc Natl Acad Sci USA* 103(45):16623–16633.

3. Pentelute BL, et al. (2008) X-ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers. *J Am Chem Soc* 130(30):9695–9701.

4. Davies PL (2014) Ice-binding proteins: A remarkable diversity of structures for stopping and starting ice growth. *Trends Biochem Sci* 39(11):548–555.

5. Todde G, Whitman C, Hovmöller S, Laaksonen A (2014) Induced ice melting by the snow flea antifreeze protein from molecular dynamics simulations. *J Phys Chem B* 118(47):13527–13534.

6. Lin FH, Graham LA, Campbell RL, Davies PL (2007) Structural modeling of snow flea antifreeze protein. *Biophys J* 92(5):1717–1723.

7. Creighton T (1993) *Proteins: Structures and Molecular Properties* (Freeman, New York).

8. Willard L, et al. (2003) VADAR: A Web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res* 31(13):3316–3319.

9. Chan HS, Dill KA (1997) Solvation: How to obtain microscopic energies from partitioning and solvation experiments. *Annu Rev Biophys Biomol Struct* 26:425–459.

10. Sharp KA, Nicholls A, Fine RF, Honig B (1991) Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science* 252(5002):106–109.

11. Chan HS, Bromberg S, Dill KA (1995) Models of cooperativity in protein folding. *Philos Trans R Soc Lond B Biol Sci* 348(1323):61–70.

12. Pentelute BL, Gates ZP, Dashnau JL, Vanderkooi JM, Kent SB (2008) Mirror image forms of snow flea antifreeze protein prepared by total chemical synthesis have identical antifreeze activities. *J Am Chem Soc* 130(30):9702–9707.

13. Jha AK, Colubri A, Freed KF, Sosnick TR (2005) Statistical coil model of the unfolded state: Resolving the reconciliation problem. *Proc Natl Acad Sci USA* 102(37):13099–13104.

14. Jha AK, et al. (2005) Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. *Biochemistry* 44(28):9691–9702.

15. Toal S, Schweitzer-Stenner R (2014) Local order in the unfolded state: Conformational biases and nearest neighbor interactions. *Biomolecules* 4(3):725–773.

16. Shi Z, Chen K, Liu Z, Sosnick TR, Kallenbach NR (2006) PII structure in the model peptides for unfolded proteins: Studies on ubiquitin fragments and several alanine-rich peptides containing QQQ, SSS, FFF, and VVV. *Proteins* 63(2):312–321.

17. Schmid FX, Baldwin RL (1978) Acid catalysis of the formation of the slow-folding species of RNase A: Evidence that the reaction is proline isomerization. *Proc Natl Acad Sci USA* 75(10):4764–4768.

18. Huyghues-Despointes BMP, Scholtz JM, Pace CN (1999) Protein conformational stabilities can be determined from hydrogen exchange rates. *Nat Struct Biol* 6(10):910–912.

19. Jackson SE, Fersht AR (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* 30(43):10428–10435.

20. Sánchez IE, Kiefhaber T (2003) Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J Mol Biol* 325(2):367–376.

21. Bachmann A, Kiefhaber T (2001) Apparent two-state tendamistat folding is a sequential process along a defined route. *J Mol Biol* 306(2):375–386.

22. Myers JK, Pace CN, Scholtz JM (1995) Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci* 4(10):2138–2148.

23. Guinn EJ, Kontur WS, Tsodikov OV, Shkel I, Record MT, Jr (2013) Probing the protein-folding mechanism using denaturant and temperature effects on rate constants. *Proc Natl Acad Sci USA* 110(42):16784–16789.

24. Ohnishi S, Kamikubo H, Onitsuka M, Kataoka M, Shortle D (2006) Conformational preference of polyglycine in solution to elongated structure. *J Am Chem Soc* 128(50):16338–16344.

25. Tran HT, Mao A, Pappu RV (2008) Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *J Am Chem Soc* 130(23):7380–7392.

26. Teufel DP, Johnson CM, Lum JK, Neuweiler H (2011) Backbone-driven collapse in unfolded protein chains. *J Mol Biol* 409(2):250–262.

27. Karandur D, Harris RC, Pettitt BM (2016) Protein collapse driven against solvation free energy without H-bonds. *Protein Sci* 25(1):103–110.

28. Holehouse AS, Garai K, Lyle N, Vitalis A, Pappu RV (2015) Quantitative assessments of the distinct contributions of polypeptide backbone amides versus side chain groups to chain expansion via chemical denaturation. *J Am Chem Soc* 137(8):2984–2995.

29. Debye P (1947) Molecular-weight determination by light scattering. *J Phys Colloid Chem* 51(1):18–32.

30. Kohn JE, et al. (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci USA* 101(34):12491–12496.

31. Tran HT, Wang X, Pappu RV (2005) Reconciling observations of sequence-specific conformational propensities with the generic polymeric behavior of denatured proteins. *Biochemistry* 44(34):11369–11380.

32. Yoo TY, et al. (2012) The folding transition state of protein L is extensive with non-native interactions (and not small and polarized). *J Mol Biol* 420(3):220–234.

33. Konuma T, et al. (2011) Time-resolved small-angle X-ray scattering study of the folding dynamics of barnase. *J Mol Biol* 405(5):1284–1294.

34. Kathuria SV, et al. (2014) Microsecond barrier-limited chain collapse observed by time-resolved FRET and SAXS. *J Mol Biol* 426(9):1980–1994.

35. Jacob J, Krantz B, Dothager RS, Thiyagarajan P, Sosnick TR (2004) Early collapse is not an obligate step in protein folding. *J Mol Biol* 338(2):369–382.

36. Baxa MC, Haddadian EJ, Jumper JM, Freed KF, Sosnick TR (2014) Loss of conformational entropy in protein folding calculated using realistic ensembles and its implications for NMR-based calculations. *Proc Natl Acad Sci USA* 111(43):15396–15401.

37. Baxa MC, Haddadian EJ, Jha AK, Freed KF, Sosnick TR (2012) Context and force field dependence of the loss of protein backbone entropy upon folding using realistic denatured and native state ensembles. *J Am Chem Soc* 134(38):15929–15936.

38. Jacobson H, Stockmayer WH (1950) Intramolecular reaction in polycondensations. 1. The theory of linear systems. *J Chem Phys* 18(12):1600–1606.

39. Frisch MT, et al. (2009) *Gaussian 09, Revision A.02* (Gaussian, Inc., Wallingford, CT).

40. Gilson MK, Honig BH (1986) The dielectric constant of a folded protein. *Biopolymers* 25(11):2097–2119.

41. Bella J, Berman HM (1996) Crystallographic evidence for C alpha-H...O=C hydrogen bonds in a collagen triple helix. *J Mol Biol* 264(4):734–742.

42. Ramachandran GN, Sasisekharan V, Ramakrishnan C (1966) Molecular structure of polyglycine II. *Biochim Biophys Acta* 112(1):168–170.

43. Chamberlain AK, Bowie JU (2002) Evaluation of C-H cdots, three dots, centered O hydrogen bonds in native and misfolded proteins. *J Mol Biol* 322(3):497–503.

44. Bartlett GJ, Choudhary A, Raines RT, Woolfson DN (2010) n→pi* interactions in proteins. *Nat Chem Biol* 6(8):615–620.

45. Karandur D, Wong KY, Pettitt BM (2014) Solubility and aggregation of Gly(5) in water. *J Phys Chem B* 118(32):9565–9572.

46. Lattman EE, Rose GD (1993) Protein folding—what's the question? *Proc Natl Acad Sci USA* 90(2):439–441.

47. Pauling L, Corey RB (1951) Configurations of polypeptide chains with favored orientations around single bonds: Two new pleated sheets. *Proc Natl Acad Sci USA* 37(11):729–740.

48. Pauling L, Corey RB, Branson HR (1951) The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci USA* 37(4):205–211.

49. Kauzmann W (1959) Some factors in the interpretation of protein denaturation. *Adv Protein Chem* 14:1–63.

50. Dill KA (1990) Dominant forces in protein folding. *Biochemistry* 29(31):7133–7155.

51. Baldwin RL (2013) The new view of hydrophobic free energy. *FEBS Lett* 587(8):1062–1066.

52. Makhatadze GI, Privalov PL (1995) Energetics of protein structure. *Adv Protein Chem* 47:307–425.

53. Berezovsky IN, Chen WW, Choi PJ, Shakhnovich EI (2005) Entropic stabilization of proteins and its proteomic consequences. *PLoS Comput Biol* 1(4):e47.

54. Gokce I, Woody RW, Anderluh G, Lakey JH (2005) Single peptide bonds exhibit poly (pro)II ("random coil") circular dichroism spectra. *J Am Chem Soc* 127(27):9700–9701.

55. Woody RW (2009) Circular dichroism spectrum of peptides in the poly(Pro)II conformation. *J Am Chem Soc* 131(23):8234–8245.

56. Mezei M, Fleming PJ, Srinivasan R, Rose GD (2004) Polyproline II helix is the preferred conformation for unfolded polyalanine in water. *Proteins* 55(3):502–507.

57. Graf J, Nguyen PH, Stock G, Schwalbe H (2007) Structure and dynamics of the homologous series of alanine peptides: A joint molecular dynamics/NMR study. *J Am Chem Soc* 129(5):1179–1189.

58. Matthews BW, Nicholson H, Becktel WJ (1987) Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc Natl Acad Sci USA* 84(19):6663–6667.

59. Ting D, et al. (2010) Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical Dirichlet process model. *PLoS Comput Biol* 6(4):e1000763.

60. Chan HS (2000) Modeling protein density of states: Additive hydrophobic effects are insufficient for calorimetric two-state cooperativity. *Proteins* 40(4):543–571.

61. Chan HS, Zhang Z, Wallin S, Liu Z (2011) Cooperativity, local-nonlocal coupling, and nonnative interactions: Principles of protein folding from coarse-grained models. *Annu Rev Phys Chem* 62:301–326.

62. Jewett AI, Pande VS, Plaxco KW (2003) Cooperativity, smooth energy landscapes and the origins of topology-dependent protein folding rates. *J Mol Biol* 326(1):247–253.

63. Badasyan A, Liu Z, Chan HS (2008) Probing possible downhill folding: Native contact topology likely places a significant constraint on the folding cooperativity of proteins with approximately 40 residues. *J Mol Biol* 384(2):512–530.

64. Shimizu S, Chan HS (2002) Anti-cooperativity and cooperativity in hydrophobic interactions: Three-body free energy landscapes and comparison with implicit-solvent potential functions for proteins. *Proteins* 48(1):15–30.

65. Liu Z, Chan HS (2005) Desolvation is a likely origin of robust enthalpic barriers to protein folding. *J Mol Biol* 349(4):872–889.

66. Chan HS, Shimizu S, Kaya H (2004) Cooperativity principles in protein folding. *Methods Enzymol* 380:350–379.

67. Plaxco KW, Simons KT, Baker D (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 277(4):985–994.

68. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Comput Chem* 79(2):926–935.

69. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781–1802.

70. Best RB, et al. (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. *J Chem Theory Comput* 8(9):3257–3273.

71. Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. *J Comput Chem* 30(10):1545–1614.

72. Cossi MBV, Cammi R, Tomasi J (1996) Ab initio study of solvated molecules: A new implementation of the polarizable continuum model. *Chem Phys Lett* 255(4):327–335.

73. Fox NK, Brenner SE, Chandonia JM (2014) SCOPe: Structural Classification of Proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Res* 42(Database issue):D304–D309.