Protein Science (1996), 5:1584–1593. Cambridge University Press. Printed in the USA.

What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties

MARY MUNSON,[1] SUGANTHI BALASUBRAMANIAN,[1] KAREN G. FLEMING,[1] ATHENA D. NAGI,[1] RONAN O'BRIEN,[2] JULIAN M. STURTEVANT,[1,2] AND LYNNE REGAN[1]

[1] Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520
[2] Department of Chemistry, Yale University, New Haven, Connecticut 06520

(RECEIVED March 22, 1996; ACCEPTED May 24, 1996)

Abstract

Here we describe how the systematic redesign of a protein's hydrophobic core alters its structure and stability. We have repacked the hydrophobic core of the four-helix-bundle protein, Rop, with altered packing patterns and various side chain shapes and sizes. Several designs reproduce the structure and native-like properties of the wild-type, while increasing the thermal stability. Other designs, either with similar sizes but different shapes, or with decreased sizes of the packing residues, destabilize the protein. Finally, overpacking the core with larger side chains causes a loss of native-like structure. These results allow us to further define the roles of tight residue packing and the burial of hydrophobic surface area in the construction of native-like proteins.

Keywords: hydrophobic core; molecular packing; molten globule; protein design

What makes a protein a protein and distinguishes it from a collapsed polymer? A pressing goal of protein engineering and design is to understand the features that give rise to the physical properties associated with a stable, "native-like" protein structure. A well-developed secondary structure is essential and the intrinsic requirements for α-helix and β-sheet formation are becoming increasingly clear (Regan, 1994; Bryson et al., 1995). Although important for stability, the turn and loop connections between elements of secondary structure often do not play a principal role in specifying the structure and properties of a protein (Brunet et al., 1993; Castagnoli et al., 1994; Predki et al., 1996) (A.D. Nagi & L. Regan, in prep). By contrast, the packing of residues in the hydrophobic core appears to be extremely important for the structure, stability, and native-like properties of natural proteins (Kellis et al., 1989; Shortle et al., 1990; Sandberg & Terwilliger, 1991; Eriksson et al., 1992; Baldwin et al., 1993; Harbury et al., 1993; Jackson et al., 1993; Richards & Lim, 1993; Dill et al., 1995), although the details are still not well understood.

A number of reports have shown that several *de novo* designed proteins apparently reproduce the desired overall fold, but nevertheless do not manifest all of the properties that are associated with natural proteins (Regan & DeGrado, 1988; Hecht et al., 1990; Beauregard et al., 1991; Raleigh & DeGrado, 1992; Kamtekar et al., 1993; O'Shea et al., 1993; Quinn et al., 1994; Tanaka et al., 1994; Yan & Erickson, 1994). These designed proteins have been described somewhat imprecisely as more "molten-globule"-like in character. This term is used in reference to certain compact intermediates populated transiently during protein folding or at equilibrium under conditions of low pH, or with apo-proteins (Ohgushi & Wada, 1983; Haynie & Freire, 1993; Fink, 1995). Such molten globule intermediates differ from native proteins in that, although they have high levels of secondary structure, their tertiary structure, particularly the packing in the hydrophobic core, is not well defined. Consequently, their physical properties are different from those of natural or "native-like" proteins (Kuwajima, 1989; Yutani et al., 1992). Specifically, equilibrium molten globules have been shown to be significantly less stable than native proteins and display shallow or non-existent thermal denaturation transitions, suggesting little or no enthalpic component to their stability. Their chemical denaturation transitions may also show low cooperativity. In addition, the lack of well-defined tertiary structure often results in poor chemical shift dispersion of their NMR spectra and in rapid exchange rates of backbone amides with solvent (Hughson et al., 1990; Jeng et al., 1990; Chyan et al., 1993).

In the context of such considerations, the aim of the present work was to investigate systematically the degree to which the internal packing of a protein can be changed before the structure and native-like properties are disrupted. Our studies focus on the four-helix-bundle protein Rop, the natural function of which is to bind a specific RNA complex and thereby regulate

Reprint requests to: Lynne Regan, Department of Molecular Biophysics and Biochemistry, 266 Whitney Avenue, New Haven, Connecticut 06520-8114; e-mail: regan%HHVMS8@venus.cis.yale.edu.

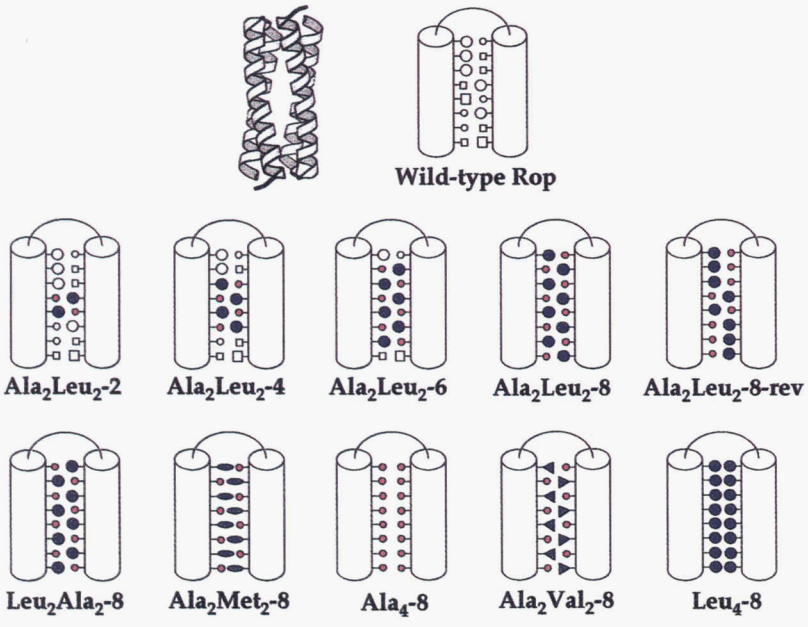

Fig. 1. Ribbon diagram (Kraulis, 1991) of wild-type Rop (Banner et al., 1987) and schematic representations showing the hydrophobic core side chains in one monomer of wild-type and the repacked mutants. Large circles represent leucine; small circles, alanine; large squares represent large, non-leucine side chains; and small squares represent small, non-alanine side chains. Ellipses represent methionine residues and triangles represent valine residues. Colored side chains indicate the repacked layers. Because Rop is an antiparallel dimer, residues in the upper layers of one monomer interact with residues in the lower layers of the other monomer and vice versa.

the replication of ColE1 plasmids (Cesareni & Banner, 1985). Rop is a homodimer of two 63-residue helix-turn-helix monomers (Banner et al., 1987) (Fig. 1). The hydrophobic core of Rop can be considered as eight layers of side chains perpendicular to the long axis of the bundle. In the "heptad repeat" nomenclature describing associating helices (Cohen & Parry, 1986), each layer is formed by two "a" and two "d" residues (Fig. 2). In most of the layers of the wild-type protein, the "a" residues are relatively small and the "d" residues relatively large, with a large residue from one layer packing against a small residue in the preceding and succeeding layers (see schematic in Fig. 1).

In an earlier study, we demonstrated that it is possible to repack the hydrophobic core of Rop using exclusively alanine at the "a" positions and leucine at the "d" positions to generate stable proteins with native-like structures and properties (Munson et al., 1994a). In the present study, we extend these findings by investigating how the stability and structural properties of Rop can be modulated by varying either the number of repacked alanine and leucine layers or the pattern and identity of the hydrophobic residues at the "a" and "d" positions.

Results and discussion

Hydrophobic core designs

A set of mutants was designed in which either two, four, six, or eight of the layers of the hydrophobic core were replaced by layers containing alanine at the "a" positions and leucine at the "d" positions (Fig. 1). Each of the mutants is named according to the identity of the residues at the "a" and "d" positions of the repacked layer and the number of repacked layers; for example, Ala_2Leu_2-8 has all eight layers repacked with alanine in the "a" positions and leucine in the "d" positions (see Materials and methods). As can be seen in the schematic in Figure 1, the wild-type Rop has a "reversal" in the packing in the second and seventh layers: a large residue occupies the "a" position, and a small residue, the "d" position. This reversed layer may cause a slight curvature at the ends of the bundle and this "end effect" may be important for tight, specific RNA binding. To explore the importance of the reversed layer, we reversed the second and seventh layers in the Ala_2Leu_2-8 mutant to make the packing more similar to wild-type; these two layers contain leucines in the "a" positions and alanines in the "d" positions (called Ala_2Leu_2-8-rev) (see Fig. 1). Furthermore, to address the possible influence of this end effect on the properties of the other

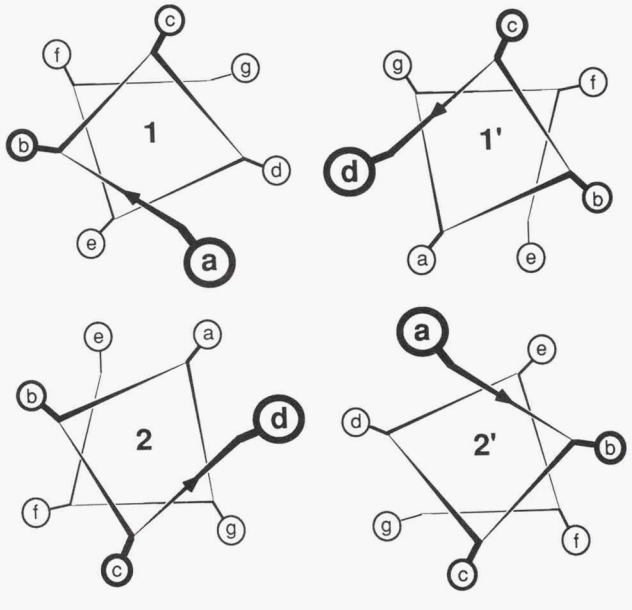

Fig. 2. Heptad repeat diagram of two layers of Rop showing the positions of the "a" and "d" residues, which pack between the four helices to form the hydrophobic core. Helices 1 and 2 are from one monomer; helices 1′ and 2′ are from the antiparallel second monomer. Arrows indicate the direction of the peptide chain, in the N-terminal to C-terminal direction.

repacked mutants, all the mutants discussed below were created in both the six- and eight-layer versions.

In addition to the Ala_2Leu_2-based designs, we created proteins in which all the alanines and leucines at the "a" and "d" positions are reversed, such that the amino acid composition of each layer is maintained, but the packing within and between layers is changed (Leu_2Ala_2). Other repacked proteins were created with the "a" positions, alanine, but with the "d" position residue varied (Ala_2Met_2 and Ala_2Ile_2). Note that the total volume of the buried core residues in wild-type Rop is 1,060 Å^3 per monomer and is very similar, within 20Å^3, in Leu_2Ala_2, Ala_2Ile_2, and Ala_2Met_2 (Chothia, 1975); the estimated buried hydrophobic surface area per monomer of wild-type is 2,300 Å^2, and the Leu_2Ala_2, Ala_2Ile_2, and Ala_2Met_2 are within 90 Å^2 (less than one alanine residue) of the wild-type (Chothia, 1976). However, the packing interactions within and between layers differ substantially from wild-type due to the pattern of packing (Leu_2Ala_2) and the side chain shapes (Ala_2Ile_2 and Ala_2Met_2). Finally, designs containing layers in which the sum of side chain volumes is either significantly decreased (Ala_2Val_2 and Ala_4) or increased (Leu_4) were constructed to determine the effects of underpacking and overpacking the core (Fig. 1). Ala_2Val_2 and Ala_4 have core volumes per monomer of 940 and 740 Å^3, respectively, and have estimated buried surface areas of 2,160 Å^2 and 1,840 Å^2, respectively. The Leu_4 has a total volume of core residues per monomer of 1,350 Å^3 and an estimated buried surface area of 2,720 Å^2.

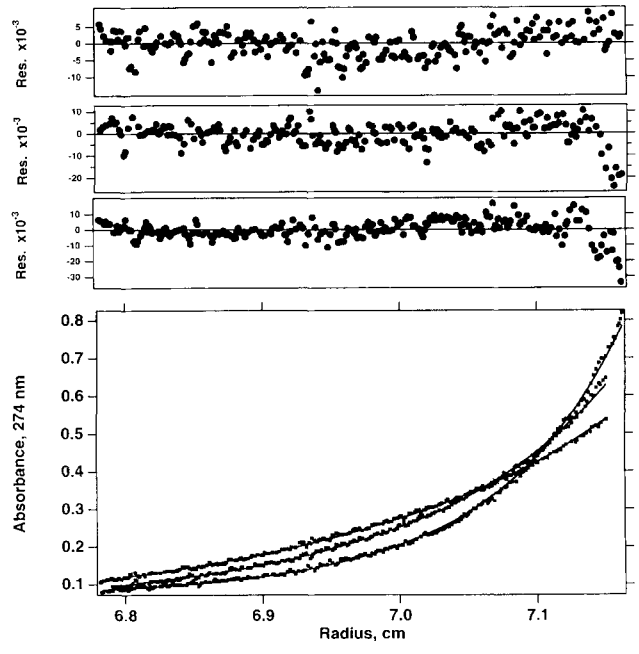

Fig. 3. Representative sedimentation equilibrium data set for Leu_2Ala_2-8 at rotor speeds 20,000 rpm (top curve); 24,500 rpm (middle curve); and 30,000 rpm (bottom curve). Solid lines represent least-squares fits to a monomer/dimer/tetramer model. Corresponding residuals are shown in the upper panels. We estimate that the dissociation constant for the monomer-dimer equilibrium for the wild-type, Ala_2Leu_2-8-rev, Leu_2Ala_2-8, and Ala_2Ile_2-6 proteins is ≤ 1 μM, and that for the dimer-tetramer equilibrium is ≥ 5 mM. The dissociation constant for the monomer-dimer equilibrium for Ala_2Met_2-8 is estimated to be 10–100-fold higher than that of the other proteins and it also has the same tendency to form tetramers at μM concentrations. The Leu_4-8 protein was found to predominantly associate as a tetramer, even at μM concentrations.

Determination of stability and structural properties

Genes encoding the repacked proteins were synthesized and the proteins overproduced in *Escherichia coli*, purified, and their properties characterized (Munson et al., 1994a, 1994b). The oligomerization state of the proteins was determined by analytical ultracentrifugation (Fig. 3) and their CD spectra were compared, taking the mean residue ellipticity (MRE) at 222 nm as an indicator of overall helicity (Table 1). The similarity of the structure of the repacked proteins to that of the wild-type was assessed by determining their ability to bind to Rop's RNA substrate using a gel mobility shift assay (Table 1). The side chains involved directly in RNA binding reside on one helix from each monomer: helix 1 and its symmetry related partner helix 1′ (Predki et al., 1995). Specific, high affinity binding requires the correct positioning of five residues from each helix, all of which are distributed along the length of the bundle. It is not known what magnitude of structural perturbation is required to abolish RNA binding. Even a slight change such as a rotation of the helices, which would leave the structure of the four-helix-bundle intact, might completely disrupt binding. However, if a mutant does bind RNA with an affinity comparable to that of the wild-type protein, we can say that the structure is very similar to that of the wild-type. Finally, the 1D ^{1}H NMR spectra of the proteins were compared (Fig. 4).

The thermodynamic properties of the proteins were assessed and compared with those of the wild-type by monitoring their thermally induced denaturation transitions by both CD and differential scanning calorimetry (DSC), and their chemically induced denaturation transitions by CD (Fig. 5; Tables 1, 2). Also, the bulk ^{1}H-^{2}H exchange rate of backbone amide protons was measured for selected mutants (Fig. 4). Amide exchange studies allow us to calculate protection factors relative to an unstructured peptide and provide a measure of both the stability and conformational dynamics of the structure (Bai et al., 1993).

The aim of these characterizations was to determine not only how stability changes as the internal packing is varied, but also to what extent certain packing arrangements generate proteins with physical properties that differ from those of typical proteins. For clarity, other than for the Ala_2Leu_2 family, we only show data for the eight-layer version of the repacked proteins. In all cases, the behavior of the six- and eight-layer versions was qualitatively similar, with the eight-layer version showing slightly higher stability. The properties of the repacked proteins can be discussed conveniently in terms of four distinct classes.

Class I

This class is comprised of the two-, four-, six-, and eight-layer Ala_2Leu_2 proteins, and also Ala_2Leu_2-8-rev (Fig. 1). The CD spectra of all proteins in this class are similar to that of the wild-type, and they are distinguished by their ability to recognize and bind to Rop's RNA substrate with affinities comparable to that of the wild-type protein (Table 1). The only protein with reduced affinity is Ala_2Leu_2-8, whose dissociation constant is increased about five fold relative to wild-type (Munson et al., 1994a). We believe that RNA binding is compromised in Ala_2Leu_2-8 because it lacks the "reversed" end layer discussed above. This interpre-

Table 1. *Summary of thermodynamic stability and activity*

Rop mutant	MRE[a] (degree cm^2 $dmol^{-1}$)	T_m[b] (°C)	$\Delta G°$[c] (kcal mol^{-1})	C_m[d] (M)	m[e] (kcal mol^{-1} M^{-1})	Binds RNA[f]
Wild-type	30,100	64	−7.7	3.3	2.4	Y
Ala_2Leu_2-2	29,200	72	−7.7	2.5	3.2	Y
Ala_2Leu_2-4	28,300	68	−5.8	2.5	2.3	Y
Ala_2Leu_2-6	27,300	82	−8.1	2.7	3.1	Y
Ala_2Leu_2-8	28,500	91	−7.5	3.0	2.8	Y
Ala_2Leu_2-8-rev	27,500	91	−9.9	3.3	3.1	Y
Leu_2Ala_2-8	30,500	−[g]	−12.8	4.1	3.2	N
Ala_2Met_2-8	28,600	48	−3.1	0.73	4.9	N
Ala_4-8	<4,800	<2	ND[h]	ND[h]	ND[h]	ND[h]
Ala_2Val_2-8	<3,400	ND[h]	ND[h]	ND[h]	ND[h]	ND[h]
Leu_4-8	25,700	−[g]	NA[i]	6.7	4.6	N

[a] MRE is the mean residue ellipticity recorded at 222 nm at 25 °C, except for Ala_4-8 and Ala_2Val_2-8, which were recorded at 2 °C.
[b] T_m is the melting temperature from the CD thermal denaturation studies.
[c] $\Delta G°$ is the extrapolated Gibbs free energy in the absence of denaturant.
[d] C_m is the concentration of denaturant at the midpoint of the denaturation transition.
[e] The m value is the slope of the line when ΔG is plotted against concentration of GuHCl.
[f] Y, yes, RNA binding observed with an affinity within a factor of five to wild-type. N, no RNA binding detectable.
[g] −, no transition observable within the temperature range accessible by CD.
[h] ND, not determined.
[i] NA, not applicable. The high stability and lack of upper baseline precluded accurate calculation of ΔG for Leu_4-8. Estimating the upper baseline would predict a ΔG greater than 30 kcal mol^{-1}.

tation is supported by the observation that Ala_2Leu_2-8-rev binds RNA with an affinity equal to that of the wild-type protein. Therefore, apart from this end effect, all the Class I proteins have structures that are very similar to the wild-type. Ultracentrifugation studies on Ala_2Leu_2-8-rev, as a representative of this class, confirm a predominantly dimeric structure (see legend of Fig. 3).

The 1D ^1H NMR spectra of the two most thermally stable mutants in this class, Ala_2Leu_2-8 and Ala_2Leu_2-8-rev, were compared (Fig. 4). The line widths, chemical shift dispersion, and appearance of the aromatic/amide region of the spectrum downfield of 6 ppm are comparable to that of the wild-type. In addition, 2D COSY and TOCSY experiments on a fully D_2O exchanged sample of Ala_2Leu_2-8-rev showed a single set of cross-peaks corresponding to F14 and Y49, confirming a symmetrical dimeric structure (data not shown).

All five of the Ala_2Leu_2 mutants have elevated melting temperatures with respect to that of the wild-type (Tables 1, 2). In general, as the number of Ala_2Leu_2 layers increases, the melting temperature increases, with a 27 °C difference between the most stable Ala_2Leu_2 proteins and the wild-type. All Class I proteins show cooperative thermal denaturation transitions with substantial associated denaturation enthalpies clearly evident from the DSC studies; results for Ala_2Leu_2-8-rev, the most thermally stable representative of this class, are shown in Figure 5A and B and Table 2.

When the stabilities of the Ala_2Leu_2 proteins toward chemical denaturation by GuHCl are compared, the stability increases as

Table 2. *Summary of differential scanning calorimetry*[a]

Rop mutant	$t_{1/2}$[b] (°C)	ΔH_{cal}[c] (kcal mol^{-1})	$\Delta H_{vH}/\Delta H_{cal}$[d]	ΔC_p[e] (kcal K^{-1} mol^{-1})	$\Delta\Delta G$(73.9 °C)[f] (kcal mol^{-1})
Wild-type	73.9 ± 0.8	109.8 ± 6.4	0.60 ± 0.04	0.56 ± 0.39	0
Ala_2Leu_2-8-rev	94.3 ± 1.4	103.6 ± 4.4	0.50 ± 0.04	1.47 ± 0.72	+4.9
Leu_2Ala_2-8	114.5 ± 0.6	125.7 ± 5.1	0.58 ± 0.10	(−0.03)	+12.5
Ala_2Met_2-8	49.2 ± 1.7	94.8 ± 10.2	0.51 ± 0.08	0.27 ± 0.41	−7.5

[a] Values listed are mean values ± the standard error of the mean.
[b] $t_{1/2}$ is the temperature of half-completion of the DSC thermal denaturation transition.
[c] ΔH_{cal} is the calorimetric enthalpy.
[d] $\Delta H_{vH}/\Delta H_{cal}$ is the ratio of the van't Hoff enthalpy to the calorimetric enthalpy. This should be 0.5 for a two-state transition of a dimeric protein.
[e] ΔC_p is the change in heat capacity that accompanies the thermal denaturation.
[f] $\Delta\Delta G(73.9\,°C) = \Delta G°_u$ (mutant) − $\Delta G°_u$ (wild-type), at the value of $t_{1/2}$ for the wild-type protein (73.9 °C).

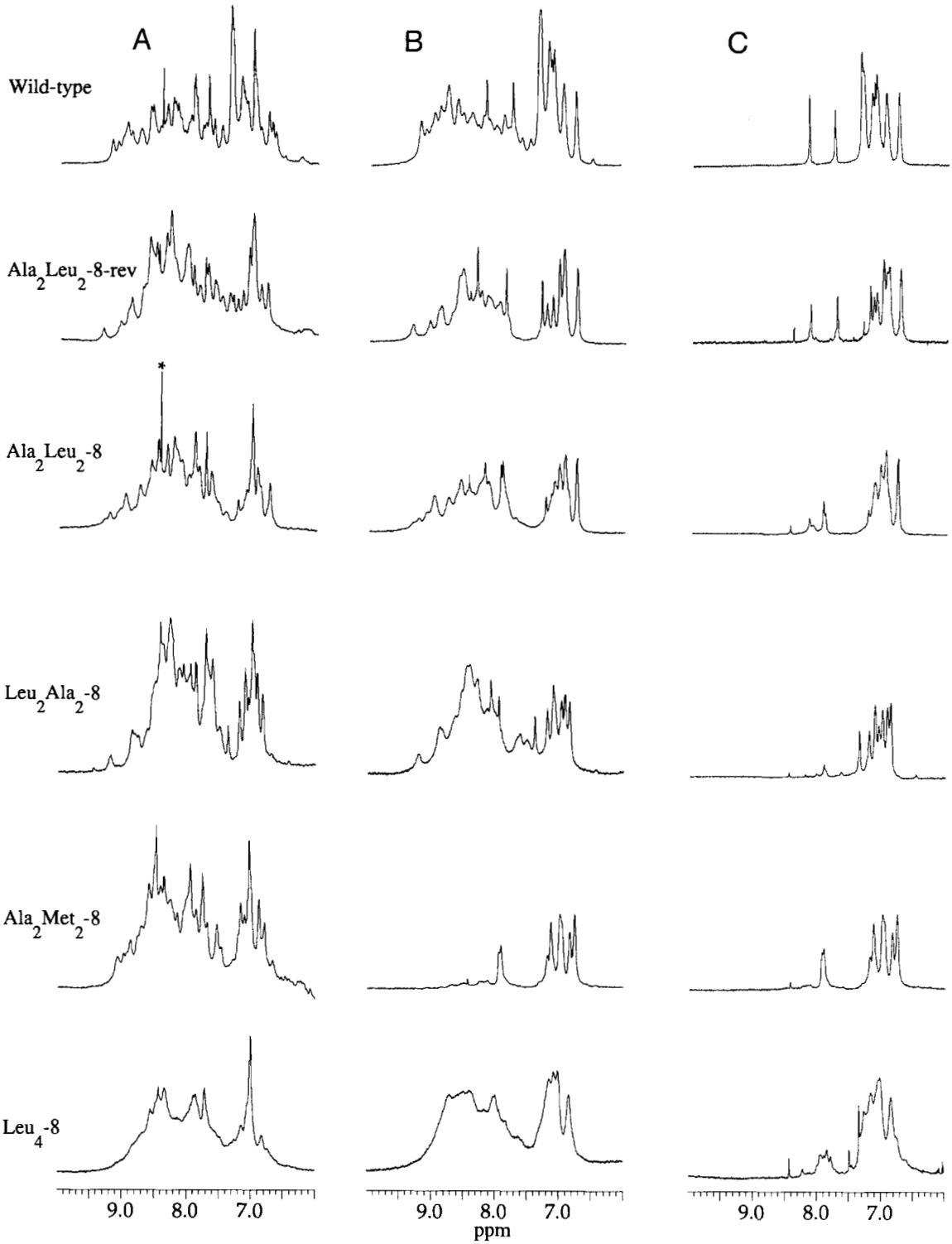

Fig. 4. 1D ^{1}H-NMR of the wild-type and selected mutants. **A:** H_2O spectra. **B:** First spectra after D_2O exchange was initiated. **C:** Fully exchanged D_2O spectra. The first and last D_2O spectra are plotted at the same scale, normalized with respect to a nonexchanging aromatic resonance. The H_2O spectra cannot be scaled in this way and are plotted to approximately the same maximum peak heights. * denotes a contaminant peak, likely formate (R.O. Fox, pers. comm.).

the number of repacked layers increases (Table 1; Fig. 5C). The stabilities of Ala_2Leu_2-2, -4, -6, and -8 are, however, slightly lower than that of the wild-type; considering the increased thermal stabilities of these mutants, the reason for this decrease is unclear. Ala_2Leu_2-8-rev has a stability slightly greater than wild-type, presumably related to the end effect described above. It is relevant to note that the behavior of wild-type Rop in the presence of GuHCl is unusual—it takes about 48 h to completely

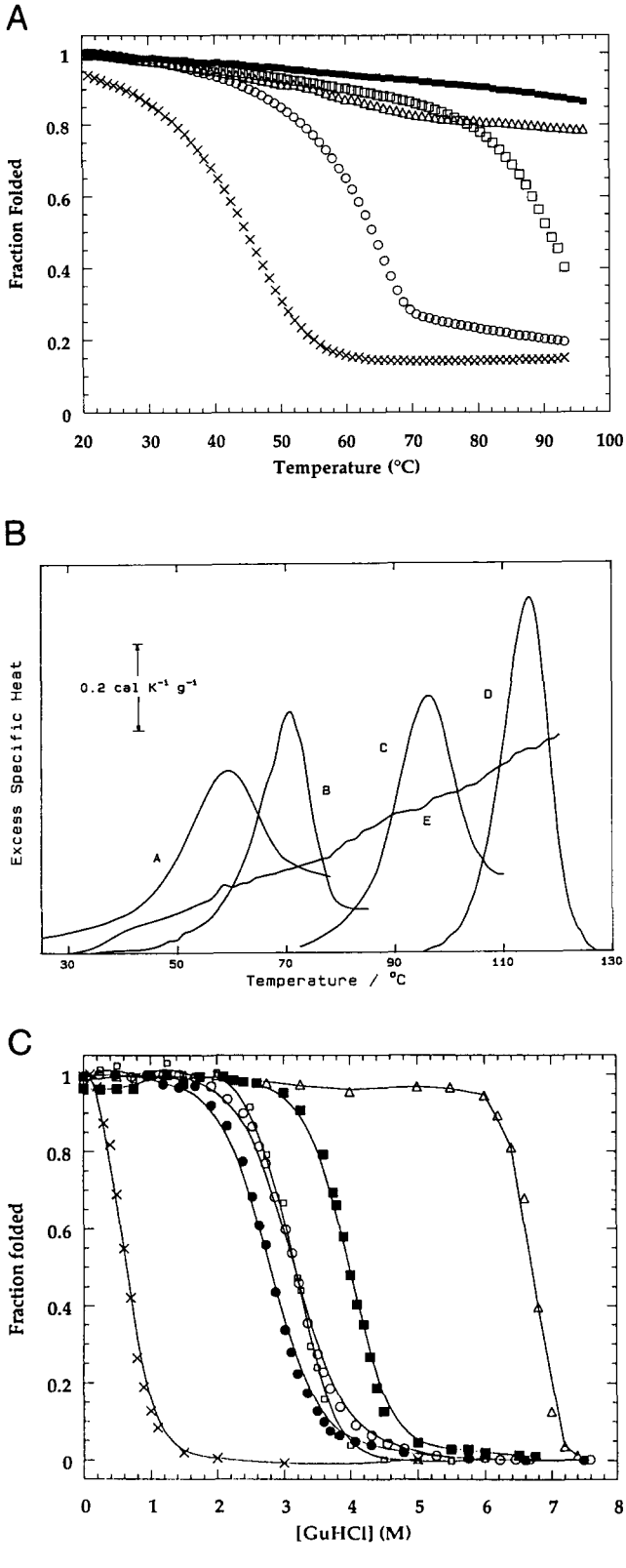

Fig. 5. **A:** Representative thermal denaturation curves monitored by CD at 222 nm for wild-type Rop (○), Ala_2Leu_2-8-rev (□), Leu_2Ala_2-8 (■), Ala_2Met_2-8 (×), and Leu_4-8 (△). **B:** Representative DSC scans for Ala_2Met_2 (A), wild-type (B), Ala_2Leu_2-8-rev (C), Leu_2Ala_2-8 (D), and Leu_4-8 (E). **C:** Representative GuHCl denaturation curves monitored by CD at 222 nm. Symbols are the same as in A, with the addition of Ala_2Leu_2-8 (●).

unfold in 4 M GuHCl (Munson et al., 1994a). By contrast, all the classes of repacked proteins fold and unfold rapidly, and, for the Class I Ala_2Leu_2 proteins, the rate increases as the number of repacked layers increases (M. Munson & L. Regan, in prep.).

The bulk rate of exchange of the backbone amide protons with solvent was determined using NMR to monitor ^1H-^2H exchange for Ala_2Leu_2-8 and Ala_2Leu_2-8-rev (Fig. 4). A fit of the exchange data to a double exponential gave a rate constant of about 1.5×10^{-5} s^{-1} for the slowest phase. At pH 7, this rate corresponds to a bulk protection factor, relative to the rate of exchange in an unstructured peptide, of approximately 2×10^4, a value that is comparable to that observed for stable natural proteins (Bai et al., 1993, 1994). The exchange rate of the wild-type protein is even slower, with an estimated rate constant of 0.3×10^{-5} s^{-1}, and a protection factor of 1×10^5. The measured exchange rate provides an indication of the stability and dynamics of the protein. This bulk exchange rate, however, cannot be used to calculate the global stabilization free energy of the protein because the rate does not necessarily correspond to the slowest exchanging amide protons.

The properties of the Class I proteins demonstrate that it is possible to use a simple, repetitive, exclusively hydrophobic repacking scheme to generate proteins that reproduce the structure, activity, and native-like properties of the wild-type. It had been suggested that, although hydrophobic clusters provide the driving force for the collapse of an unfolded protein, specific polar interactions are required to achieve a unique native fold (Behe et al., 1991; Fink, 1995). The properties of the Class I mutants demonstrate that appropriately packed hydrophobic residues can generate a unique structure that is not only native-like, but is also very thermally stable.

Class II

Class II contains the proteins Leu_2Ala_2 and Ala_2Ile_2, which differ from Class I by one key feature: they have lost the ability to bind the target RNA complex. Therefore, although their CD spectra are indicative of a high helicity comparable to that of the wild-type, their structures must differ to some degree. Ultracentrifugation studies confirm that both Class II proteins are predominantly dimers (see legend of Fig. 3). The thermodynamic properties of the Leu_2Ala_2 protein are discussed below. Ala_2Ile_2 has very similar properties, to be discussed in detail elsewhere (M.A. Willis & L. Regan, in prep.).

The ^1H NMR spectrum of Leu_2Ala_2-8 shows good chemical shift dispersion, and again 2D COSY and TOCSY experiments confirm the presence of a symmetrical dimer (Fig. 4). The melting temperature of Leu_2Ala_2-8 is remarkably high (Fig. 5A,B) and yet a cooperative denaturation transition can be detected, under pressure, with a $t_{1/2}$ of 115 °C and a significant associated enthalpy (Table 2). The protein is also stabilized toward denaturation by GuHCl, with a midpoint of 4.1 M (Fig. 5C). The rate of amide proton exchange is slow, with an estimated rate constant of 0.7×10^{-5} s^{-1}, corresponding to a protection factor of 4×10^4. This slightly slower rate of exchange, compared with that of Ala_2Leu_2-8 and Ala_2Leu_2-8-rev, is consistent with its higher thermal and chemical stability.

The properties of the Class II mutants suggest that, although the Leu_2Ala_2 and Ala_2Ile_2 proteins have structures that differ from wild-type, both these arrangements are able to pack well

within the context of the four-helix-bundle to generate extremely stable proteins with native-like physical characteristics.

Class III

The Class III mutants, Ala_2Met_2, Ala_2Val_2, and Ala_4, not only lack the ability to bind Rop's RNA substrate, but are all significantly less stable than wild-type (Tables 1, 2). Only the Ala_2Met_2-8 protein forms dimers and ultracentrifugation studies suggest that the dimerization of Ala_2Met_2-8 is weaker than that of wild-type (see legend of Fig. 3). The CD spectrum of Ala_2Met_2-8 is similar to that of the wild-type, with a comparable ellipticity at 222 nm. Its ^1H NMR spectrum is well dispersed, and again the number of aromatics visible in the fully exchanged spectrum is consistent with the presence of a symmetrical dimer (Fig. 4). The melting temperature of Ala_2Met_2-8 is reduced by 16 °C relative to wild-type, and the midpoint of its GuHCl-induced chemical denaturation is 2.6 M less than wild-type (Fig. 5); however, Ala_2Met_2-8 still displays cooperative denaturation transitions. The amide protons of Ala_2Met_2-8 exchange extremely rapidly at pH 7; exchange is essentially complete by the first time point measured. The fast exchange observed, in comparison with the Class I and II proteins, reflects the significantly reduced stability and perhaps increased conformational dynamics of Ala_2Met_2-8. These observations suggest that the presence of methionine in the "d" helical positions results in poor hydrophobic core packing; however, this packing is sufficient to stabilize a weakly associating dimer. Additionally, Ala_2Met_2-8 retains native-like properties, demonstrating that poor packing alone does not necessarily cause molten-like behavior.

The two other members of Class III are the "underpacked" Ala_2Val_2 and Ala_4 proteins. Both these proteins are predominantly unfolded. At 2 °C, the CD spectrum of Ala_4-8 is indicative of some helical structure. This helicity can be thermally denatured with a concentration independent broad denaturation transition, suggesting that the observed helicity derives from a monomeric species (data not shown) (Marqusee et al., 1989). Presumably, although Ala_4-8 cannot bury sufficient hydrophobic surface area or volume to stabilize a dimeric structure, the high α-helix-forming propensity of alanine (Bryson et al., 1995) gives rise to some helical character. With Ala_2Val_2-8, the combination of insufficient buried hydrophobic volume and the low intrinsic α-helix forming propensity of valine (Bryson et al., 1995) results in an unfolded protein.

Class IV

The Leu_4-8 mutant of Class IV has a considerably "overpacked" hydrophobic core. The structural and thermodynamic consequences of this overpacking are dramatic. The Leu_4-8 protein does not bind RNA, yet it has an MRE that is indicative of high helical content (Table 1). A combination of equilibrium and velocity sedimentation ultracentrifugation studies suggest that the predominant oligomerization state of Leu_4-8 is a tetramer, rather than a dimer (see Fig. 3 legend and Materials and methods). The ^1H NMR spectrum of Leu_4-8 shows some chemical shift dispersion, but the spectra are qualitatively poorer than those of wild-type and the other mutants studied (Fig. 4). This is probably caused by a combination of effects: the degeneracy due to the presence of 36 leucines, a perturbation of the symmetry of the protein, and the higher molecular weight of the tetramer. The appearance of the aromatic region of the fully D_2O exchanged protein is significantly different from that of the wild-type and the other repacked proteins.

The thermodynamic properties of this protein are remarkable. No cooperative thermal denaturation transition is observed by either CD or DSC (Fig. 5A,B). A gradual loss of less than 20% of total helicity is observed between approximately 40 °C and 80 °C when the transition is monitored by CD, and when monitored by DSC, a gradual increase in heat capacity is observed over the temperature range studied. It could be that the protein is extremely resistant to denaturation by heat, and we are unable to reach the temperature needed to induce a cooperative denaturation transition. Alternatively, the protein may never undergo a cooperative thermal denaturation transition, but structure may continue to be lost in the same gradual fashion if it could be taken to higher temperatures.

Leu_4-8 is also extremely resistant to chemical denaturation (Fig. 5C; Table 1). The midpoint of the GuHCl-induced transition is 6.8 M, a concentration more than 3 M higher than that of wild-type Rop. The chemical denaturation behavior of Leu_4-8 is in contrast to the thermal denaturation behavior, because not only is a transition observed, but it is extremely cooperative.

The rate of bulk amide exchange for Leu_4-8 is somewhat faster than that observed for the wild-type. It has an estimated rate constant of $0.5 \times 10^{-5}\ s^{-1}$, corresponding to a protection factor of about 6×10^4, which is comparable to that observed for the most thermally stable repacked mutant, Leu_2Ala_2-8. The amide exchange properties of Leu_4-8 are similar to those of typical, stable natural proteins. By contrast, equilibrium molten globules and several designed proteins have much lower protection factors (Hughson et al., 1990; Handel et al., 1993; Davidson et al., 1995), the highest being in the range $10–10^3$.

Summary

The results we present demonstrate how the stability and structural properties of a protein can be altered dramatically by the nature of the hydrophobic core residues. Repacking the core with exclusively hydrophobic Ala_2Leu_2 layers generates stable proteins with native-like properties and structures very similar to wild-type. Alternative packing schemes can also generate well-packed cores. Leu_2Ala_2 and Ala_2Ile_2, for example, give rise to proteins that have structures different from wild-type, but that retain native-like characteristics and are extremely stable. Not all combinations of residues in the core, however, result in stable proteins. Favorable steric interactions and burial of sufficient hydrophobic volume and surface area are both important to give a stable protein. In the Ala_2Met_2 mutant, the methionine residues, which have similar volume and buried surface area to leucine, but are shaped differently, must pack poorly in the core. This packing results in a protein that is significantly less stable than the wild-type, although it still displays native-like properties. Ala_4 is unable to bury sufficient hydrophobic volume and surface area to generate a stably folded protein, whereas Ala_2Val_2 cannot bury enough to overcome the destabilization due to its poor helical propensity.

Finally, with the overpacked Leu_4-8 variant, we have created a protein with unusual properties. First, the predominant species is a tetramer, rather than a dimer. Second, the protein is extremely resistant to both thermal and chemical denaturation, and has a rate of amide exchange similar to natural proteins.

What is the origin of this extreme stability? It is unlikely that a tetramer containing 72 leucine residues will achieve the complementary tight packing associated with natural proteins and predicted for the Class I and II repacked cores. It appears, therefore, that for Leu_4-8, the increased burial of hydrophobic surface area of these leucines is sufficient to compensate for the energetic cost of poor packing. This balance of forces is significantly different from those that stabilize natural proteins, and is reflected in the physical properties of Leu_4. Studies to determine the exact arrangement and mobility of the hydrophobic core residues in Leu_4-8 and the other repacked proteins are in progress.

Materials and methods

Cloning and purification

Mutant Rop proteins were cloned, overproduced, and purified as previously described (Munson et al., 1994a, 1994b), except for the Leu_2Ala_2 mutant: this protein was produced from the high copy number plasmid pET11d (Novagen) to improve protein yields, and an additional Q sepharose column (Pharmacia) was added to the purification protocol to improve the purity of the final preparation. The RNA binding assays were performed as described previously (Munson et al., 1994a). Several proteins have been renamed in this report: wild-type was previously Rop11; Ala_2Leu_2-6 was Rop13; Ala_2Leu_2-8 was Rop21. The naming of the mutants is in the form $\mathbf{X}_2\mathbf{Y}_2$, where **X** are the "a" position residues and **Y** are the "d" position residues. In the case of the Ala_2Leu_2 mutants, the number of repacked layers of the core, starting from the center of the bundle, is given at the end of the name, i.e., Ala_2Leu_2-4 means that the middle four layers are Ala_2Leu_2. The sequence of each of the mutants is: GlyThrLysGln**X**LysThr**Y**LeuAsnMet**X**ArgPhe**Y**ArgSerGln**X** LeuThr**Y**LeuGluLys**X**AsnGlu**Y**Asp**X**AspGlu**Y**AlaAspIle**X**Glu Ser**Y**HisAspHis**X**AspGlu**Y**TyrArgSer**X**LeuAla**YX**GlyAspAsp GlyGluAsnLeu. The wild-type protein contains (in order of appearance in the linear sequence): **X** = Glu, Ala, Thr, Leu, Ala, Cys, Ala, Cys, Phe; and **Y** = Ala, Ile, Leu, Leu, Gln, Leu, Leu, Arg. Ala_2Leu_2-2 has the wild-type sequence except for Ala at residue 19 and Leu at 15. Ala_2Leu_2-4 is Ala_2Leu_2-2 with Ala at residue 38. Ala_2Leu_2-6 has **X** = Ala and **Y** = Leu, except for wild-type residues at positions 5, 55, and 56. Ala_2Leu_2-8 has **X** = Ala and **Y** = Leu. Ala_2Leu_2-8-rev has **X** = Ala and **Y** = Leu, except Ala at 8 and 34, and Leu at 26, 34, and 52. Leu_2Ala_2-8 has **X** = Leu and **Y** = Ala. Ala_2Met_2-8 has **X** = Ala and **Y** = Met. Ala_2Val_2-8 has **X** = Ala, **Y** = Val. Ala_4-8 has **X** = **Y** = Ala. Leu_4-8 has **X** = **Y** = Leu.

CD and DSC measurements

The CD measurements were taken on an AVIV 62DS instrument (Lakewood, New Jersey) and the DSC measurements were made in either a DASM-4 calorimeter (Biopribor, Puschino, Russia) or an MC2 instrument (Microcal, Inc., Northhampton, Massachusetts) as described previously (Munson et al., 1994a). All CD measurements were performed using a protein concentration of 20 μM, whereas the concentration for DSC was about 0.3 mM; this accounts for the increased $t_{1/2}$ over the CD measured T_m. Protein samples were dialyzed in 100 mM Na phosphate buffer, pH 7, 200 mM NaCl (plus 1 mM DTT for the wild-type). All the MRE values were calculated from ellipticities recorded at 25 °C, except for Ala_2Val_2 and Ala_4, which were recorded at 2 °C. The T_ms in Table 1 were estimated by calculating the temperature at which the slope of the first derivative of the CD data was a minimum (Munson et al., 1994a). GuHCl data for both Ala_2Met_2-8 and Leu_4-8 are estimated due to lack of lower and upper baselines, respectively. Values of ΔC_p evaluated from individual DSC experiments are subject to large uncertainty, and introduce uncertainties in extrapolations to evaluate $\Delta G°$ at temperatures other than $T_{1/2}$ by means of the equation:

$$\Delta G_T^\circ = \Delta H_{T1/2}\,(1 - T/T_{1/2}) + \Delta C_p[T - T_{1/2} + T\ln(T/T_{1/2})].$$

For the mutants, we therefore report values of $\Delta\Delta G$, relative to the $\Delta G°$ of wild-type Rop, at 73.9 °C. The reported values for GuHCl and the thermal stabilities for the wild-type protein (Rop11) and for the GuHCl stabilities of the Ala_2Leu_2-6 (Rop13) and Ala_2Leu_2-8 (Rop21) proteins in Tables 1 and 2 are slightly different than reported previously (Munson et al., 1994a) due to more accurate calculations for the GuHCl experiments and averaging of additional DSC experiments.

Analytical ultracentrifugation

Ultracentrifugation was performed using a Beckman Optima XL-A analytical ultracentrifuge at 25 °C with protein samples in 50 mM Na phosphate buffer, pH 7.0, 150 mM NaCl; the buffer for wild-type also contained 2 mM β-mercaptoethanol. The partial specific volumes of all proteins, and estimations of hydration (for the sedimentation velocity calculations) were calculated based on the amino acid composition (Cohn & Edsall, 1943; Kunz & Kauzmann, 1974). Equilibrium data were collected at multiple protein concentrations and multiple speeds (ranging from 20,000 to 40,000 rpm) using two sector cells. A minimum of three data sets for each protein were subjected to global analysis using the MacNONLIN algorithm (Johnson et al., 1981). Goodness of fit was determined by examination of the residuals and minimization of the variance. Association constants were estimated using a molar extinction coefficient at 278 nm of 2,840 (McRorie & Voelker, 1993).

Velocity data for the Leu_4-8 protein were collected at 60,000 rpm and evaluated by second moment analysis using Origin software (Beckman) to yield experimental weight-average values for s. These s_w values were compared with theoretical s_{max} values for equivalent hydrated spheres to determine if the measured values were consistent with the oligomeric state as predicted from the equilibrium experiment. The weight-average s value obtained was greater than the theoretical maximum s value calculated for the dimer and was consistent with a tetramer.

NMR

Spectra were obtained at 10 °C on a Varian Unity 500 MHz spectrometer. The proton chemical shifts were referenced with respect to the residual water signal assumed to resonate at 4.932 ppm at 10 °C. Water suppression of the spectra recorded in 90% H_2O/10% D_2O was achieved by presaturation of the water peak. To initiate the amide proton exchange experiments, proteins lyophilized from H_2O containing 100 mM Na phosphate buffer, pH 7.0, and 200 mM NaCl were dissolved in D_2O. The

minimum dead-time between addition of D_2O to the lyophilized protein and the start of data acquisition was 3 min. Multiple 1D spectra were recorded over a period of 20 h using a spectral width of 8,000 Hz in 16,000 data points with a recycle delay of 1.5 s. All the spectra were processed using the program Felix 2.3 (Biosym Technologies) with a 1 Hz exponential line-broadening window function. At each time point, the total NH intensities were determined by integration over the downfield region of the spectrum from approximately 7.5 ppm to 9.5 ppm. Intensities were normalized with respect to nonexchanging aromatic resonances, the intensities of which remained invariant throughout the exchange experiment. In the absence of any sequence-specific assignments, the bulk amide exchange behavior of the different mutants were compared to estimate protection factors. It is not possible to fit the bulk exchange kinetics to a single exponential because it is dependent on the decay behavior of many individual amide protons. The curves fit well, however, to a sum of nonlinear exponentials of the form $I = I_1 e^{-k1t} + I_2 e^{-k2t}$, where I is the total intensity observed at time t, k_1 and k_2 are rate constants for each decay phase, and I_1 and I_2 are their respective amplitudes. The quality of the fits were assessed by visual inspection of fits and their χ^2 and R values. Protection factors were estimated as the ratio of k_{int}/k_{ex}, where k_{int} is the intrinsic amide exchange rate constant at the appropriate pH and temperature (Bai et al., 1993) and k_{ex} is the observed amide exchange rate constant. The overall random coil exchange rates were obtained by summing the predicted exchange behavior of all the individual backbone amides and fitting the resultant exchange curve to a mono-exponential decay (Bai et al., 1993). The number of amides in the D_2O spectra were estimated based on the area under the peak corresponding to the assigned Y49 ring resonances from 2D COSY and TOCSY experiments. Based on these estimates, of a total of 63 backbone amides, the slow phase amplitudes correspond to approximately 31 protons per monomer for the wild-type, 18 protons for Ala_2Leu_2-8-rev, 15 protons for Ala_2Leu_2-8, and 44 for Leu_2Ala_2-8. Unambiguous assignment of the aromatic resonances in Leu_4-8 was not possible; hence, we have not attempted to determine precisely the number of protons in the D_2O spectrum. From our exchange analysis, we estimate that about 50% of the protons decay with an apparent rate constant of 0.5×10^{-5} s^{-1}.

Acknowledgments

L.R. is an NSF National Young Investigator and a Dreyfus Teacher Scholar. This work was supported by NIH grant GM49146-01A1. K.G.F. is a NIH NRSA Postdoctoral Fellow (GM16769). We thank Preston Hensley of SmithKline Beecham Pharmaceutical, Inc. (King of Prussia, Pennsylvania) for helpful discussions; Boehringer Ingelheim Pharmaceutical, Inc. (Ridgefield, Connecticut), and Pfizer Central Research (Groton, Connecticut) for providing access to their ultracentrifuges. We are grateful to Fred Richards for thoughtful discussions and members of the Regan laboratory for comments on the manuscript. We also thank Jenny Yang for advice on the amide exchange experiments.

References

Bai Y, Milne JS, Mayne L, Englander SW. 1993. Primary structure effects on peptide group hydrogen exchange. *Proteins Struct Funct Genet 17*:75–86.

Bai Y, Milne JS, Mayne L, Englander SW. 1994. Protein stability parameters measured by hydrogen exchange. *Proteins Struct Funct Genet 20*:4–14.

Baldwin EP, Hajiseyedjavadi O, Baase WA, Matthews BW. 1993. The role of backbone flexibility in the accommodation of variants that repack the core of T4 lysozyme. *Science 262*:1715–1718.

Banner DW, Kokkinidis M, Tsernoglou D. 1987. Structure of the ColE1 Rop protein at 1.7 Å resolution. *J Mol Biol 196*:657–675.

Beauregard M, Goraj K, Goffin V, Heremans K, Goormaghtigh E, Ruysschaert JM, Martial JA. 1991. Spectroscopic investigation of structure in Octarellin (a de novo protein designed to adopt the α/β-barrel packing). *Protein Eng 4*:745–749.

Behe MJ, Lattman EE, Rose GD. 1991. The protein-folding problem: The native fold determines packing, but does packing determine the native fold? *Proc Natl Acad Sci USA 88*:4195–4199.

Brunet AP, Huang ES, Huffine ME, Loeb JE, Weltman RJ, Hecht MH. 1993. The role of turns in the structure of an α-helical protein. *Nature 364*:355–358.

Bryson JW, Betz SF, Lu HS, Suich DJ, Zhou HX, O'Neil KT, DeGrado WF. 1995. Protein design: A hierarchic approach. *Science 270*:935–941.

Castagnoli L, Vetriani C, Cesareni G. 1994. Linking an easily detectable phenotype to the folding of a common structural motif: Selection of rare turn mutations that prevent the folding of Rop. *J Mol Biol 237*:378–387.

Cesareni G, Banner DW. 1985. Regulation of plasmid copy number by complementary RNAs. *Trends Biochem Sci 17*:303–306.

Chothia C. 1975. Structural invariants in protein folding. *Nature 254*:304–308.

Chothia C. 1976. The nature of the accessible and buried surfaces in proteins. *J Mol Biol 105*:1–14.

Chyan CL, Wormald C, Dobson CM, Evans PA, Baum J. 1993. Structure and stability of the molten globule state of Guinea-pig α-lactalbumin: A hydrogen exchange study. *Biochemistry 32*:5681–5691.

Cohen C, Parry DAD. 1986. α-Helical coiled coils—A widespread motif in proteins. *Trends Biochem Sci 11*:245–248.

Cohn EJ, Edsall JT. 1943. Density and apparent specific volume of proteins. In: *Proteins, amino acids, and peptides*. New York: Reinhold Publishing, Inc. pp 370–381.

Davidson AR, Lumb KJ, Sauer RT. 1995. Cooperatively folded proteins in random sequence libraries. *Nature Struct Biol 2*:856–864.

Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. 1995. Principles of protein folding—A perspective from simple exact models. *Protein Sci 4*:561–602.

Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, Matthews BW. 1992. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science 255*:178–183.

Fink AL. 1995. Compact intermediate states in protein folding. *Annu Rev Biophys Biomol Struct 24*:495–522.

Handel TM, Williams SA, DeGrado WF. 1993. Metal ion-dependent modulation of the dynamics of a designed protein. *Science 261*:879–885.

Harbury PB, Zhang T, Kim PS, Alber T. 1993. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science 262*:1401–1407.

Haynie DT, Freire E. 1993. Structural energetics of the molten globule state. *Proteins Struct Funct Genet 16*:115–140.

Hecht MH, Richardson JS, Richardson DC, Ogden RC. 1990. De novo design, expression, and characterization of Felix: A four-helix bundle protein of native-like sequence. *Science 249*:884–891.

Hughson FM, Wright PE, Baldwin RL. 1990. Structural characterization of a partly folded apomyoglobin intermediate. *Science 249*:1544–1548.

Jackson SE, Moracci M, elMasry N, Johnson CM, Fersht AR. 1993. Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. *Biochemistry 32*:11259–11269.

Jeng MF, Englander SW, Elöve GA, Wand AJ, Roder H. 1990. Structural description of acid-denatured cytochrome *c* by hydrogen exchange and 2D NMR. *Biochemistry 29*:10433–10437.

Johnson M, Correia JJ, Yphantis DA, Halvorson H. 1981. Analysis of data from the analytical ultracentrifuge by nonlinear least squares techniques. *Biophys J 36*:575–588.

Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH. 1993. Protein design by binary patterning of polar and nonpolar amino acids. *Science 262*:1680–1685.

Kellis JT Jr, Nyberg K, Fersht AR. 1989. Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry 28*:4914–4922.

Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Kunz ID, Kauzmann W. 1974. Hydration of proteins and polypeptides. *Adv Prot Chem 28*:239–345.

Kuwajima K. 1989. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins Struct Funct Genet 6*:87–103.

Marqusee S, Robbins VH, Baldwin RL. 1989. Unusually stable helix formation in short alanine-based peptides. *Proc Natl Acad Sci USA 86*: 5286–5290.

McRorie DK, Voelker PJ. 1993. *Self-associating systems in the analytical ultracentrifuge*. Fullerton, California: Beckman Instruments, Inc. p 20.

Munson M, O'Brien R, Sturtevant JM, Regan L. 1994a. Redesigning the hydrophobic core of a four-helix-bundle protein. *Protein Sci 3*:2015–2022.

Munson M, Predki PF, Regan L. 1994b. ColE1-compatible vectors for high-level expression of cloned DNAs from the T7 promoter. *Gene 144*:59–62.

Ohgushi M, Wada A. 1983. "Molten-globule state:" A compact form of globular proteins with mobile side-chains. *FEBS Lett 164*:21–24.

O'Shea EK, Lumb KJ, Kim PS. 1993. Peptide "Velcro*": Design of a heterodimeric coiled coil. *Curr Biol 3*:658–667.

Predki PF, Agrawal V, Brünger AT, Regan L. 1996. Amino-acid substitutions in a surface turn modulate protein stability. *Nature Struct Biol 3*:54–58.

Predki PF, Nayak LM, Gottlieb MB, Regan L. 1995. Dissecting RNA-protein interactions: RNA-RNA recognition by Rop. *Cell 80*:41–50.

Quinn TP, Tweedy NB, Williams RW, Richardson JS, Richardson DC. 1994. Betadoublet: De novo design, synthesis, and characterization of a β-sandwich protein. *Proc Natl Acad Sci USA 91*:8747–8751.

Raleigh DP, DeGrado WF. 1992. A de novo designed protein shows a thermally induced transition from a native to a molten globule-like state. *J Am Chem Soc 114*:10079–10081.

Regan L. 1994. Structural biology: Born to be beta. *Curr Biol 7*:656–658.

Regan L, DeGrado WF. 1988. Characterization of a helical protein designed from first principles. *Science 241*:976–978.

Richards FM, Lim W. 1993. An analysis of packing in the protein folding problem. *Q Rev Biophys 26*:423–498.

Sandberg W, Terwilliger T. 1991. Energetics of repacking a protein interior. *Proc Natl Acad Sci USA 88*:1706–1710.

Shortle D, Stites W, Meeker A. 1990. Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. *Biochemistry 29*:8033–8041.

Tanaka T, Kuroda Y, Kimura H, Kidokoro SI, Nakamura H. 1994. Cooperative deformation of a de novo designed protein. *Protein Eng 7*:969–976.

Yan Y, Erickson BW. 1994. Engineering of Betabellin 14D: Disulfide-induced folding of a β-sheet protein. *Protein Sci 3*:1069–1073.

Yutani K, Ogasahara K, Kuwajima K. 1992. Absence of the thermal transition in apo-α-lactalbumin in the molten globule state. *J Mol Biol 228*: 347–350.