Speeding up protein folding: mutations that increase the rate at which Rop folds and unfolds by over four orders of magnitude

Mary Munson[1,2], Karen S Anderson[3] and Lynne Regan[1]

Background: The dimeric four-helix-bundle protein Rop folds and unfolds extremely slowly. To understand the molecular basis for the slow kinetics, we have studied the folding and unfolding of wild-type Rop and a series of hydrophobic core mutants.

Results: Mutation of the hydrophobic core creates stable, dimeric, and wild-type-like proteins with dramatically increased rates of both folding and unfolding. The increases in rates are dependent upon the number and position of repacked residues within the hydrophobic core.

Conclusions: Rop folds by a rapid collision of monomers to form a dimeric intermediate with substantial helical content, followed by a slow rearrangement to the final native structure. Rop unfolding is a single extremely slow kinetic phase. The slow steps of both folding and unfolding are dramatically increased by hydrophobic core replacements, suggesting that their main effect is to substantially decrease the energy of the transition state.

Addresses: [1]Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, New Haven, CT 06520, USA. [2]Present address: Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA. [3]Department of Pharmacology, Yale University, 333 Cedar Street, New Haven, CT 06510, USA.

Correspondence: Lynne Regan
e-mail: regan%hhvms8@venus.cis.yale.edu

Key words: circular dichroism, fluorescence, folding kinetics, protein folding, stopped-flow

Received: 11 Oct 1996
Revisions requested: 21 Nov 1996
Revisions received: 13 Dec 1996
Accepted: 18 Dec 1996

Published: 28 Jan 1997
Electronic identifier: 1359-0278-002-00077

Folding & Design 28 Jan 1997, 2:77–87

Introduction

When a protein folds, the backbone and sidechain atoms organize from the extensive number of conformations present in the unfolded state to the limited number populated in the folded state. Because protein folding usually occurs on the order of milliseconds to seconds, it is generally accepted that folding must proceed through a particular pathway or set of pathways, rather than being a random process [1,2].

The folding and unfolding of small monomeric proteins, including the α domain of Trp synthase, lysozyme, ribonuclease, chymotrypsin inhibitor 2, barnase, and cytochrome *c*, have been the focus of many elegant studies. These proteins generally fold and unfold quickly (milliseconds or faster) and many aspects of their folding pathways have been elucidated [3–9]. Small dimeric proteins, such as the *trp* aporepressor, P22 Arc repressor, and small leucine zippers, have also been well studied and have been found to fold and unfold at rates that range from milliseconds to minutes [10–15].

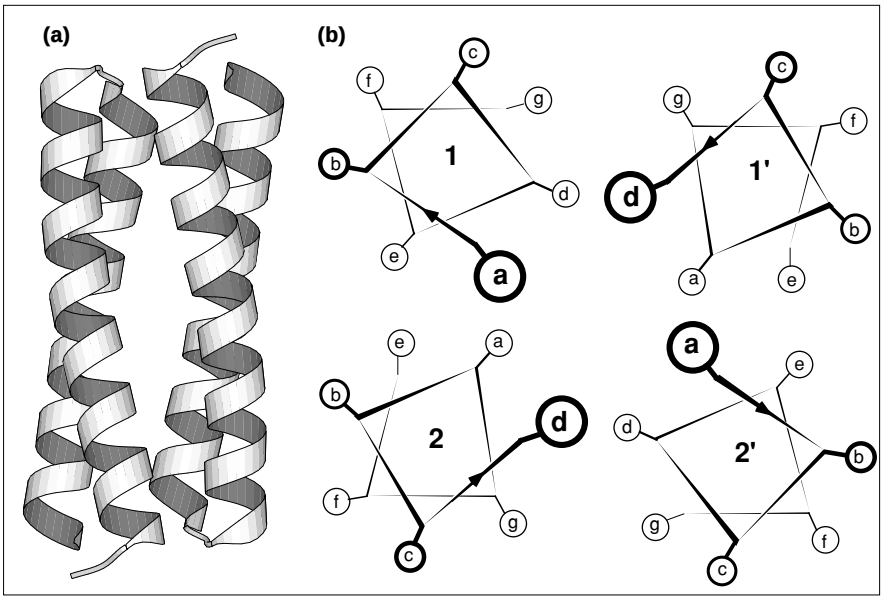
An important goal in the study of protein folding is to understand the molecular basis for the differences in magnitude of the kinetic folding and unfolding barriers that are observed for different proteins. A few specific examples of high energy barriers have been well characterized. The *cis*/*trans* isomerization of certain proline residues can cause slow folding kinetics, as can disulfide bond formation [5]. Recently, it has been suggested that the formation of buried polar interactions may also slow the rate at which a protein folds [14]. Studies that focus on defining the nature of additional rate-limiting steps are essential for a comprehensive description of protein folding.

Rop is a dimeric four-helix-bundle protein whose function in *Escherichia coli* is to regulate ColE1 plasmid replication by binding to a complex of two complementary RNAs [16,17]. The symmetrical four-helix-bundle structure is formed by the association of two 63-residue helix-loop-helix monomers (Fig. 1a). Because Rop is a small protein that does not contain proline residues, disulfide bonds, or co-factors, one might have predicted that its folding and unfolding kinetics would be rapid. They are not. Wild-type Rop folds and unfolds extremely slowly.

In earlier studies, we described the equilibrium thermodynamic properties of a series of mutants in which the hydrophobic core of Rop was repacked using combinations of alanine and leucine residues [18,19]. Interestingly, we observed that the fully repacked Ala_2Leu_2-8 mutant (described below) does not display the slow unfolding and refolding kinetics that are characteristic of wild-type Rop. The rates of both folding and unfolding are dramatically increased in Ala_2Leu_2-8; both occur on the millisecond timescale.

To understand the molecular basis for the increases in the folding and unfolding rates that result from repacking of the Rop hydrophobic core, we studied the folding and

Figure 1



Rop structure. (a) Ribbon diagram of Rop created using the program Molscript [29] from the coordinates of the crystal structure of the wild-type protein [26]. The short, unstructured C-terminal tail is not shown. (b) Schematic diagram illustrating the heptad repeat nomenclature for associating helices and the positions of the 'a' and 'd' residues that form the Rop hydrophobic core.

unfolding of wild-type Rop together with a series of repacked mutants. The results of these studies allow us to propose a model for the folding pathway of wild-type Rop, to identify the rate-determining step, and to quantify the kinetic effects of the mutations.

Results and discussion

Wild-type Rop: folding kinetics

The equilibrium thermodynamic properties of wild-type Rop, a stable dimer with a dissociation constant of less than 1 μM, have been well characterized and are summarized in Table 1 [18,19].

For our kinetic studies, protein folding was monitored by following the change in fluorescence intensity of Tyr49. Protein was first equilibrated in a high concentration of guanidine hydrochloride (GdnHCl) until unfolding was complete. To initiate folding, the unfolded protein was diluted 10-fold into a series of GdnHCl concentrations that provided strongly folding conditions (>90% folded). When wild-type Rop is refolded to low concentrations of GdnHCl (≤1.75 M), two kinetic phases are clearly visible: a fast phase with a half-life of <<1 s that accounts for about two-thirds of the amplitude, and a slow phase with a half-life of several seconds that accounts for the remaining one-third of the amplitude (Fig. 2a). At higher concentrations of GdnHCl (>1.75 M), only the slow phase is detected (Fig. 2b).

At low protein concentrations (6–25 μM), the fast phase is linearly dependent on protein concentration, but at higher protein concentrations it becomes independent of protein concentration. The slow phase is independent of protein concentration at all concentrations tested (6–200 μM). The rate of the slow step is inversely dependent on the concentration of GdnHCl (Fig. 3a).

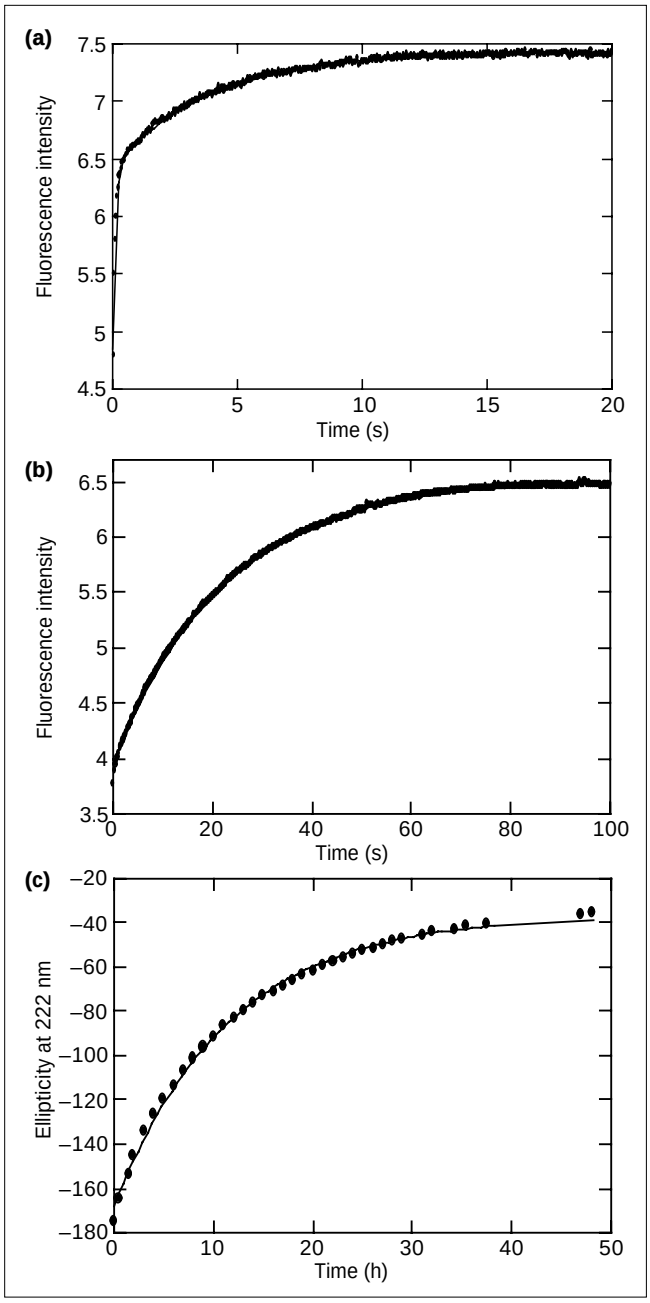
Table 1

Stability of wild-type Rop and the repacked mutants.

Rop variant	T_m (°C)	ΔG^0 (kcal mol^{-1})	C_m (M)	m (kcal mol^{-1} M^{-1})
Wild type	64	–7.7	3.3	2.4
Ala_2Leu_2-4	68	–5.8	2.5	2.3
Ala_2Leu_2-2	72	–7.7	2.5	3.2
Ala_2Leu_2-(3+6)	72	–8.4	2.7	3.3
Leu_2Ala_2-(2+7)	85	–12.8	3.6	3.7
Ala_2Leu_2-6-rev	85	–10.3	3.4	3.1
Ala_2Leu_2-8-rev	91	–9.9	3.3	3.1
Ala_2Leu_2-(2+7)	85	–8.7	2.7	3.4
Ala_2Leu_2-(1+8)	54	–6.3	2.3	2.9
Ala_2Leu_2-6	82	–8.1	2.7	2.7
Ala_2Leu_2-8	91	–7.5	3.0	2.8

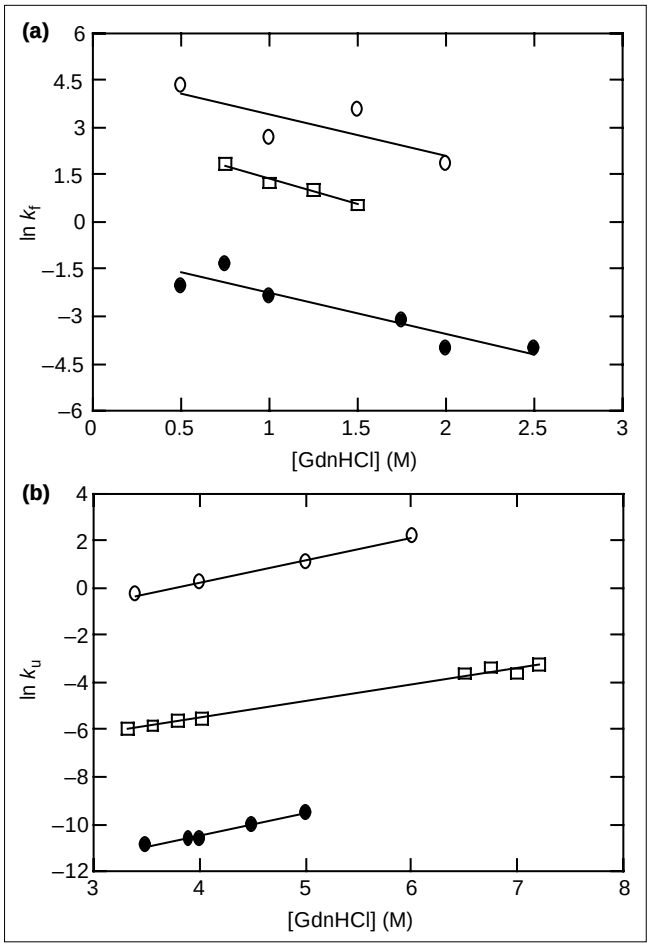
T_m is the melting temperature from the CD thermal denaturation studies, estimated by calculating the temperature at which the slope of the first derivative of the denaturation curve is a minimum. ΔG^0 is the extrapolated Gibbs free energy in the absence of denaturant, obtained from a linear plot (with R > 0.995) of ΔG versus [GdnHCl]. We estimate that the systematic error associated with such analyses is on the order of 10%. C_m is the concentration of denaturant at the midpoint of the GdnHCl-induced denaturation transition. The m value is the slope of the line when ΔG is plotted against concentration of GdnHCl. The data on wild-type Rop, Ala_2Leu_2-2, Ala_2Leu_2-4, Ala_2Leu_2-6, Ala_2Leu_2-8 and Ala_2Leu_2-8-rev are taken from [19].

Figure 2



Examples of folding and unfolding raw data for wild-type Rop. Fluorescence intensity (arbitrary units) is shown as a function of time in (a) and (b), ellipticity at 222 nm is shown as a function of time in (c). (a) Refolding from 6.2 M to 0.75 M GdnHCl; the data are fit to a double exponential. (b) Refolding from 6.2 M to 1.75 M GdnHCl; the data are fit to a single exponential. (c) Unfolding from 0.0 M to 4.0 M GdnHCl; the data are fit to a single exponential.

Figure 3

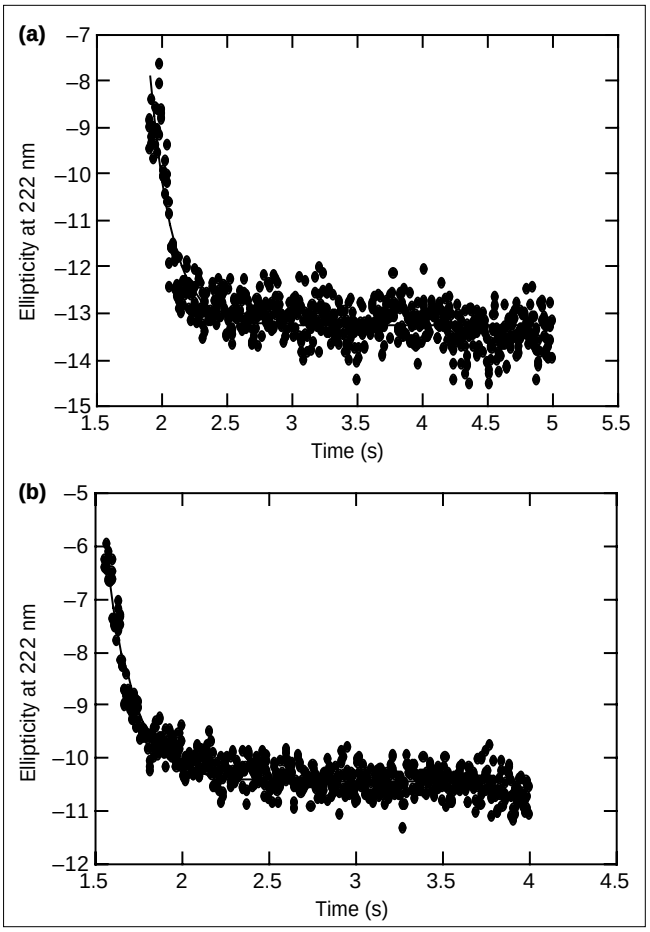


(a) Representative plots of lnk_f as a function of [GdnHCl] for wild-type Rop (closed circles), Ala_2Leu_2-(1+8) (squares), and Ala_2Leu_2-6 (open circles). (b) Representative plots of lnk_u as a function of [GdnHCl] for wild-type Rop (closed circles), Ala_2Leu_2-(1+8) (squares), and Ala_2Leu_2-6 (open circles).

We used stopped-flow circular dichroism (CD) to compare the kinetics of secondary structure formation with the kinetics of tertiary structure formation that are monitored in the fluorescence studies. It is evident from the CD data that helix formation is rapid. Acquisition of ellipticity at 222 nm occurs at a rate that is consistent with helix formation during the fast phase (Fig. 4a).

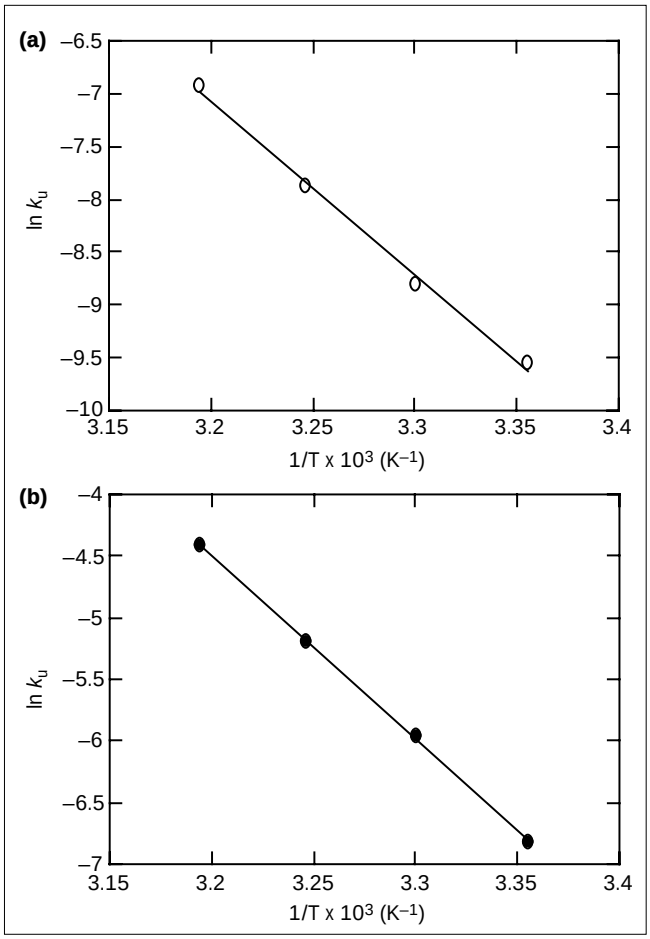
On the basis of these observations, we propose that the fast phase corresponds to the bimolecular collision and partial folding of Rop monomers to form a dimeric intermediate. This intermediate is detectable at low concentrations of GdnHCl, but at higher concentrations of GdnHCl, it is unstable and not detected as a discrete intermediate. At low protein concentrations, the rate of this fast phase is linearly dependent on protein concentration; the bimolecular collision is rate-determining under these conditions. At higher protein concentrations, the rate of the fast step is independent of protein concentration, suggesting that a unimolecular step is now rate-limiting. The fast phase is followed by a slow phase in which the final rearrangement

Figure 4



The folding of (a) wild-type Rop and (b) Ala_2Leu_2-4 monitored by stopped-flow CD. Ellipticity at 222 nm is shown as a function of time.

Figure 5



(a) Arrhenius plot for wild-type Rop unfolding. ln*k*u is plotted as a function of temperature. The unfolding reaction was performed in 5.0 M GdnHCl and monitored by following the ellipticity at 222 nm by CD. (b) Arrhenius plot for Ala_2Leu_2-4 unfolding. ln*k*u is plotted as a function of temperature. The unfolding reaction was performed in 4.0 M GdnHCl and monitored by following the ellipticity at 222 nm by CD.

of the hydrophobic core of the helical dimeric intermediate occurs and GdnHCl is excluded. The simplest model that is consistent with this data is shown in the scheme:

$$2U \leftrightarrow I_2 \leftrightarrow N_2$$

U represents unfolded monomers, which collide and interact to form a helical dimeric intermediate, I_2. I_2 rearranges slowly to form N_2, the native dimer.

Wild-type Rop: unfolding kinetics

The rate of unfolding of wild-type Rop was monitored using either CD or fluorescence. The two techniques gave identical results. Folded protein was diluted 10-fold into strongly unfolding conditions using concentrations of GdnHCl that correspond to 90% or greater unfolded protein. A single slow kinetic phase was observed that is well fit by a single exponential with an amplitude that accounts for the entire denaturation transition (Fig. 2c). Under these strongly unfolding conditions, the dimeric intermediate is too unstable to be detected; unfolding is essentially a simultaneous dissociation of the dimer and unfolding of the helices. The unfolding rate (k_u) is independent of protein concentration, as expected for a unimolecular reaction, and the plot of lnk_u against the concentration of GdnHCl is linear (Fig. 3b).

To further explore the energetics of the unfolding transition, unfolding of the wild-type protein in 5 M GdnHCl was performed at a series of different temperatures. From an Arrhenius analysis of this data, we calculate an activation energy (E_a) of 32 kcal mol^{-1} (Fig. 5a). Considering that this measurement was made in strongly unfolding conditions, the energy barrier is extraordinarily high and demonstrates why the unfolding rate is so slow for wild-type Rop. For comparison, typical activation energies that

have been measured for small proteins, in the absence of denaturant, are on the order of 15–30 kcal mol^{-1} [11,14,20–23].

Because the slow unfolding of wild-type Rop is so unusual, it was important to demonstrate that this effect is not a GdnHCl-specific phenomenon. We therefore compared the relative unfolding rates of wild-type Rop and the repacked Ala_2Leu_2-8 mutant in the presence of a different denaturant, urea. At concentrations of urea that correspond to strongly unfolding conditions, it takes over 24 h for wild-type Rop to completely unfold, whereas Ala_2Leu_2-8 is completely unfolded in less than 30 s. We also monitored the rate of unfolding in the absence of denaturant using the formation of heterodimers as an assay. Wild-type Rop is mixed with a Rop variant that has a single point mutation in the unstructured short tail that extends beyond the C terminus of helix 2. The point mutation introduces a charge difference, which causes the tail mutant to migrate differently from wild-type Rop on native gels. Formation of heterodimers, which presumably involves the dissociation and unfolding of monomers followed by association and refolding, is detected by the appearance of a third species of intermediate mobility. At 37°C, the heterodimerization reaction for wild-type Rop takes over 24 h [24], illustrating that unfolding also occurs extremely slowly in the absence of denaturant.

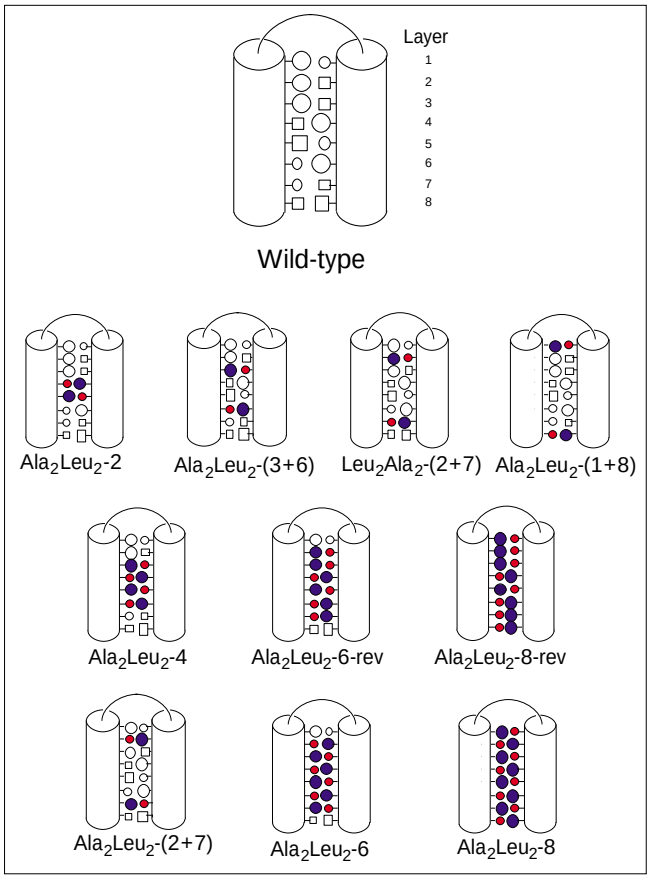
Although there are no disulfide bonds in wild-type Rop, there are two free cysteine residues per monomer. A final concern, therefore, was to be certain that the slow folding kinetics are not an artefact associated with the formation of nonnative disulfides in a folding intermediate. We discounted this possibility by showing that the presence of a large excess of a reducing agent, either dithiothreitol (DTT) or Tris(2-carboxyethyl) phosphine [25], does not change the unfolding rate. Also, a mutant in which both cysteine residues have been mutated to alanine unfolds only about 10-fold faster than the wild type. This rate enhancement is simply accounted for by the effect of repacking those layers, as discussed below (data not shown).

Having characterized the folding and unfolding of wild-type Rop, we proceeded to investigate in detail how a series of hydrophobic core redesigns perturb the observed kinetics.

Description of the hydrophobic core redesigns

The hydrophobic core of Rop is formed by eight layers of sidechains; each layer is composed of two 'a' and two 'd' position residues (Fig. 1b). The layers of wild-type Rop tend to have small residues in the 'a' positions, and large residues in the 'd' positions. The small and large residues pack against each other between adjacent layers. The packing in layers 2 and 7 is 'reversed' from this pattern; the 'a' residues are large and the 'd' residues are small (Fig. 6).

Figure 6



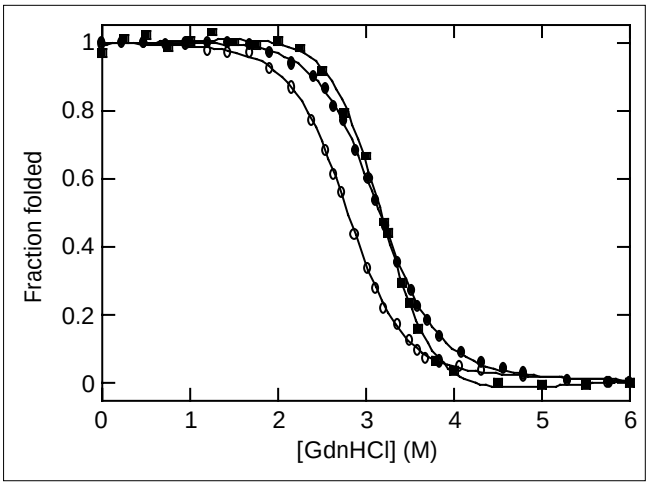
Schematic diagram of a single monomer of wild-type Rop and the repacked mutants showing the pattern of sidechains in each core. The numbering of the layers is indicated on the wild type. Small circles represent alanines and large circles represent leucines. Small squares represent non-alanine sidechains which pack as small residues in wild-type Rop, e.g. cysteine and threonine; large squares represent non-leucine sidechains which pack as large residues, e.g. isoleucine and glutamine. Colored sidechains illustrate the repacked layers.

In an earlier study, we observed that a completely repacked Ala_2Leu_2-8 mutant both folded and unfolded extremely rapidly [18]. We therefore designed a series of repacked mutants in which we varied both the number and position of the repacked layers to investigate this phenomenon.

The repacked mutants that are the focus of this study contain alanines in the 'a' positions and leucines in the 'd' positions (Ala_2Leu_2), or leucines in the 'a' positions and alanines in the 'd' positions (Leu_2Ala_2) in two or more of the eight layers of the hydrophobic core (Fig. 6). Because Rop is an antiparallel dimer, layers are repacked in equivalent pairs, e.g. layer 1 is equivalent to layer 8.

The first row of mutants in Figure 6 are those that have single pairs of layers repacked. These proteins are Ala_2Leu_2-2 (layers 4 and 5 are Ala_2Leu_2); Ala_2Leu_2-(3+6)

Figure 7



Representative GdnHCl equilibrium denaturation curves for wild-type Rop (closed circles), Ala_2Leu_2-8 (open circles), and Ala_2Leu_2-8-rev (squares). Fraction folded, determined from the ellipticity at 222 nm monitored by CD, is plotted as a function of [GdnHCl].

(layers 3 and 6 are Ala_2Leu_2); Leu_2Ala_2-(2+7) (layers 2 and 7 are Leu_2Ala_2 to reproduce the reversal seen in the wild type) and Ala_2Leu_2-(1+8) (layers 1 and 8 are Ala_2Leu_2).

The second set of mutants has increasing numbers of repacked layers. One of this set is shown on the first row of mutants in Figure 6: Ala_2Leu_2-2 (middle two layers repacked). The additional mutants with four, six or eight layers repacked are shown on the second row of mutants in Figure 6: Ala_2Leu_2-4 (middle four layers repacked), Ala_2Leu_2-6-rev (middle six layers repacked, including the reversed layers 2 and 7), and Ala_2Leu_2-8-rev (all eight layers repacked, including the reversed layers 2 and 7).

The final set of mutants was designed to explore the contribution of the reversal of packing seen in layers 2 and 7 of wild-type Rop and are shown on the third row of mutants in Figure 6. These mutants have Ala_2Leu_2 repacked layers 2 and 7 in the wild-type (Ala_2Leu_2-(2+7)), Ala_2Leu_2-6, and Ala_2Leu_2-8 backgrounds.

Structure and stability of the repacked variants

All the repacked proteins have structures that are very similar to that of the wild type. The most compelling evidence is that the repacked proteins bind Rop's target RNA complex with affinities comparable to that of wild-type Rop. In addition, the repacked proteins have mean residue ellipticities that are within 15% of that of the wild type, and the wavelength of the excitation and emission maxima of the fluorescence of the single tyrosine residue (Tyr49) are identical to that of the wild type. The susceptibility of the fluorescence to quenching by acrylamide is identical in wild type and repacked mutants, indicating similar solvent accessibility of the Tyr49 residue in all the proteins. Finally, analytical ultracentrifugation studies on the wild type and selected mutants confirm a dimeric structure with dissociation constants of less than 1 μM [19].

The thermal denaturation transitions of the wild-type and repacked proteins were monitored using CD to follow the ellipticity at 222 nm. The transitions are fully reversible and, in general, additional numbers of repacked layers increase the thermal stability of the proteins relative to wild type. The range of this effect is large: the T_m of wild-type Rop is 64°C, and that of the most stable repacked proteins, Ala_2Leu_2-8 and Ala_2Leu_2-8-rev, is 91°C. The thermal stability data are summarized in Table 1 (the proteins are listed in Table 1 in order of increasing folding rates to allow direct comparison with Table 2). Differential scanning calorimetry (DSC) on wild-type Rop and selected repacked mutants confirms that the proteins undergo a reversible two-state denaturation transition with the associated van't Hoff enthalpy equal to the calorimetric enthalpy [18,19].

GdnHCl-induced denaturation was monitored by following the change in ellipticity at 222 nm. In contrast to the dramatic increases in thermal stability, the effect of repacking on the protein's resistance to chemical denaturation is small. Most of the repacked proteins have stabilities that are similar to or are somewhat destabilized relative to the wild type; only Ala_2Leu_2-6-rev, Ala_2Leu_2-8-rev and Leu_2Ala_2-(2+7) are more stable. Representative GdnHCl-induced denaturation curves are shown in Figure 7, and the GdnHCl stability data are summarized in Table 1. The GdnHCl-induced denaturations are reversible and are well fit to a two-state transition. Identical results are obtained whether the denaturation transition is monitored by CD or by the change in the intrinsic fluorescence intensity [18,19] (A Nagi, L Regan, unpublished data). The decoupling of the protein's thermal and chemical stability is unusual and is the subject of ongoing studies. It is not addressed further in this paper, but it is probably associated with differences in the stability of the protein at their respective T_ms, compared with their stabilities at 25°C (M Mahoney, L Regan, unpublished data). These differences derive from differences in the heat capacity change (ΔCp) upon denaturation of the wild-type versus mutant proteins.

Repacked proteins: folding kinetics

The folding of the repacked mutants was monitored by following the change in fluorescence intensity of Tyr49, exactly as described for the wild type. Folding rates for all the mutant proteins were increased over that of the wild type. For mutants in which the rate of folding was least accelerated (Ala_2Leu_2-(3+6), Ala_2Leu_2-2, Ala_2Leu_2-4, and Leu_2Ala_2-(2+7)), two kinetic phases were detected. The

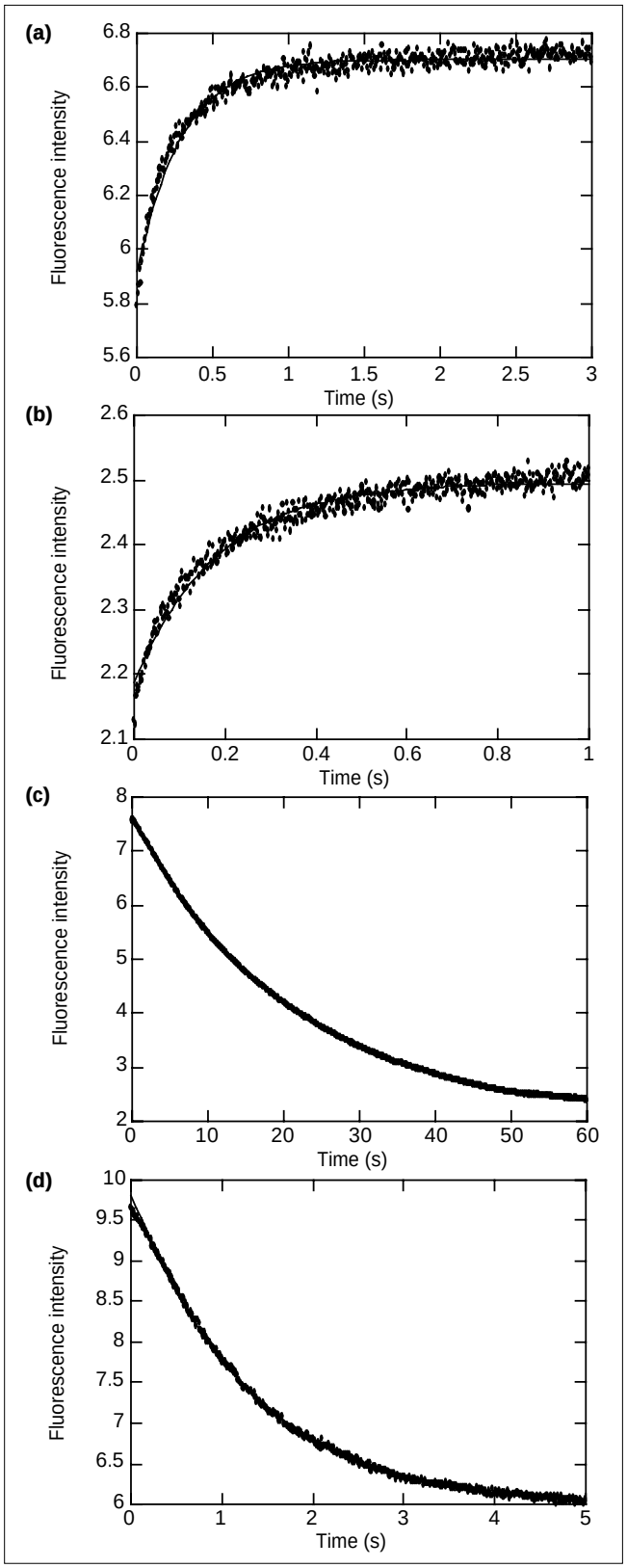


Figure 8

(a) Refolding of Ala_2Leu_2-(1+8) from 6.2 M to 1.0 M GdnHCl; the data are fit to a double exponential. (b) Refolding of Ala_2Leu_2-6 from 6.2 M to 2.0 M GdnHCl; the data are fit to a single exponential. Fluorescence intensity (arbitrary units) is plotted as a function of time. (c) Unfolding of Ala_2Leu_2-(1+8) from 0.0 M to 4.0 M GdnHCl; the data are fit to a single exponential. (d) Unfolding of Ala_2Leu_2-6 from 0.0 M to 3.4 M GdnHCl; the data are fit to a single exponential. Fluorescence intensity (arbitrary units) is plotted as a function of time.

rates of both phases were faster than the corresponding phases for wild type. For mutants in which the rate of folding was greatly increased, only a single phase was detected. The amplitude of the single phase we observe with these mutants represents less than one-quarter of the total amplitude, suggesting that a significant amount of folding (i.e. the fast phase and some of the 'slow phase') is accomplished during the dead time of the stopped-flow (approximately 5 ms). Representative examples of a 'medium' and a 'fast' mutant folding are shown in Figure 8a,b.

The limited information available on the fast phase precludes a detailed quantitative comparison of the wild type and mutants. For the slow phase, however, we were able to measure the rate of folding at a number of concentrations of GdnHCl, and to compare quantitatively the effects of repacking. For all the repacked proteins, the rate of the slow phase was inversely linearly proportional to the concentration of GdnHCl. Representative plots for the medium and fast folding mutants, Ala_2Leu_2-(1+8) and Ala_2Leu_2-6 respectively, are shown in Figure 3a, along with the wild type data. The folding rate constants for all the mutant proteins are summarized in Table 2, in order of increasing folding rate.

We again followed up on the fluorescence studies by using stopped-flow CD to monitor the kinetics of secondary structure formation. The fastest folding mutants all acquire essentially full helicity within the dead time of the instrument. For the slower folding mutants, similar to our observations with the wild type, substantial helicity is acquired rapidly, consistent with the rapid formation of a helical intermediate (Fig. 4b). These results suggest that the hydrophobic core mutations do not perturb the pathway of folding, but simply accelerate the rates of both the fast and slow folding steps.

Repacked proteins: unfolding kinetics

We measured the unfolding rate of each of the repacked mutants using fluorescence or CD, as described for the wild type. For all the repacked proteins, the unfolding kinetics were well fit to a single exponential. Representative examples are shown in Figure 8c,d. As we observed for wild-type Rop, the unfolding rates are independent of protein concentration. A plot of $\ln k_u$ versus concentration of GdnHCl is linear and examples are shown, along with the wild type, in Figure 3b. The unfolding rates of all the mutants are increased over that of wild type, and the data are summarized in Table 2.

Table 2

Kinetics of unfolding and folding.

Rop variant	$[Gu]_f$ (M)	k_f (s^{-1})	Relative k_f	$[Gu]_u$ (M)	k_u (s^{-1})	Relative k_u
Wild type	2.6	0.013	1	4.0	2.4×10^{-5}	1
Ala_2Leu_2-4	2.0	0.020	1.5	3.1	6.6×10^{-4}	28
Ala_2Leu_2-2	1.7	0.041	3.2	2.8	4.2×10^{-4}	18
Ala_2Leu_2-(3+6)	1.8	0.097	7.5	3.1	2.0×10^{-4}	8.3
Leu_2Ala_2-(2+7)	2.9	0.13	10	3.9	4.4×10^{-4}	18
Ala_2Leu_2-6-rev	2.8	1.1	85	3.9	1.6×10^{-2}	6.7×10^{2}
Ala_2Leu_2-8-rev	2.7	1.2	92	3.8	6.4×10^{-2}	2.7×10^{3}
Ala_2Leu_2-(2+7)	1.9	1.6	1.2×10^{2}	3.0	0.17*	7.1×10^{3}
Ala_2Leu_2-(1+8)	1.4	2.1	1.6×10^{2}	3.4	2.7×10^{-3}	1.1×10^{2}
Ala^2Leu_2-6	2.1	4.0	3.1×10^{2}	3.4	0.75*	3.1×10^{4}
Ala_2Leu_2-8	2.2	7.9	6.1×10^{2}	3.6	1.2*	5.0×10^{4}

The Rop proteins are listed in order of increasing folding rates. The folding and unfolding rates are compared at the same final fraction folded or unfolded (90 %). $[Gu]_f$ is the concentration of GdnHCl where the protein is 90 % folded, determined in the equilibrium experiments. k_f is the folding rate at this concentration of GdnHCl, determined by linear extrapolation or interpolation of the plot of $\ln k_f$ versus [GdnHCl]. The protein concentration is 100 μM. All values of k_f were determined by stopped-flow fluorimetry. This rate only corresponds to the slow phase of the folding reaction, which is protein concentration independent. The fast phase is detected for the wild type at low concentrations of GdnHCl, and for the Ala_2Leu_2-(3+6), Ala_2Leu_2-2, Ala_2Leu_2-4, and Leu_2Ala_2-(2+7) mutants. The relative k_f shown is relative to the wild-type folding rate: relative $k_f = k_{f(mut)}/k_{f(wild\text{-}type)}$. $[Gu]_u$ is the concentration of GdnHCl where the protein is 90 % unfolded, determined in the equilibrium experiments. k_u is the unfolding rate at this concentration of GdnHCl, determined by linear extrapolation or interpolation of the plot of $\ln k_u$ versus [GdnHCl]. The protein concentration is 20 μM. All values of k_u were determined by CD, except the values indicated by the asterisk, which were determined by stopped-flow fluorimetry at 100 μM protein. Relative k_u is relative to the wild-type unfolding rate: relative $k_u = k_{u(mut)}/k_{u(wild\text{-}type)}$.

We performed an Arrhenius analysis of the unfolding reaction on the Ala_2Leu_2-4 mutant. This protein was chosen because it has an unfolding rate that can be monitored easily by conventional CD up to 40°C in 4.0 M GdnHCl; this concentration of GdnHCl corresponds to the same final fraction folded as that used in the Arrhenius analysis of wild-type Rop. The Arrhenius plot for Ala_2Leu_2-4 (Fig. 5b) gives a calculated activation energy of unfolding of 30 kcal mol^{-1}. The 2 kcal mol^{-1} difference between this mutant and the wild type corresponds to an approximately 10^2 fold difference in the rate of unfolding.

A detailed comparison of the kinetic effects of the mutations

The most dramatic effect of the repacked mutants is to increase the rate of both the slow folding step and the slow unfolding step. Because both folding and unfolding are faster, the primary effect of the mutations must be to lower the activation barrier between the helical intermediate and the folded state (illustrated schematically in Fig. 9).

The rate of the fast folding step, i.e. the conversion of unfolded monomers to helical intermediate, is also enhanced. A combination of two general effects could account for this increase. The higher helical propensities of alanine and leucine could give rise to populations of partially folded helices with an equilibrium that is shifted toward the folded state relative to the wild-type sequence. Also, mutant proteins rich in alanine and leucine residues could more readily associate to form the intermediate through hydrophobic interactions that are not necessarily identical to those that stabilize the native protein. Because the rate of this step in many of the mutants is too fast to quantify, however, we limit our further discussion of the different mutants to their effects on the slow step.

The slow folding step involves the final rearrangement of the hydrophobic core to the native structure, with concomitant exclusion of GdnHCl. Conversely, in unfolding, the slow step corresponds to a loosening of the hydrophobic packing in the core that allows entry of GdnHCl. It is remarkable that the effect of mutations in the hydrophobic core is not 'all or nothing'; these effects depend upon the precise number and location of the repacked Ala_2Leu_2 layers and are additive. There are three general effects. First, repacking an individual layer with Ala_2Leu_2 (or Leu_2Ala_2 in layers 2 and 7) increases the rates of folding and unfolding approximately 10-fold, regardless of the location of the layer within the core (Fig. 10a). Second, the

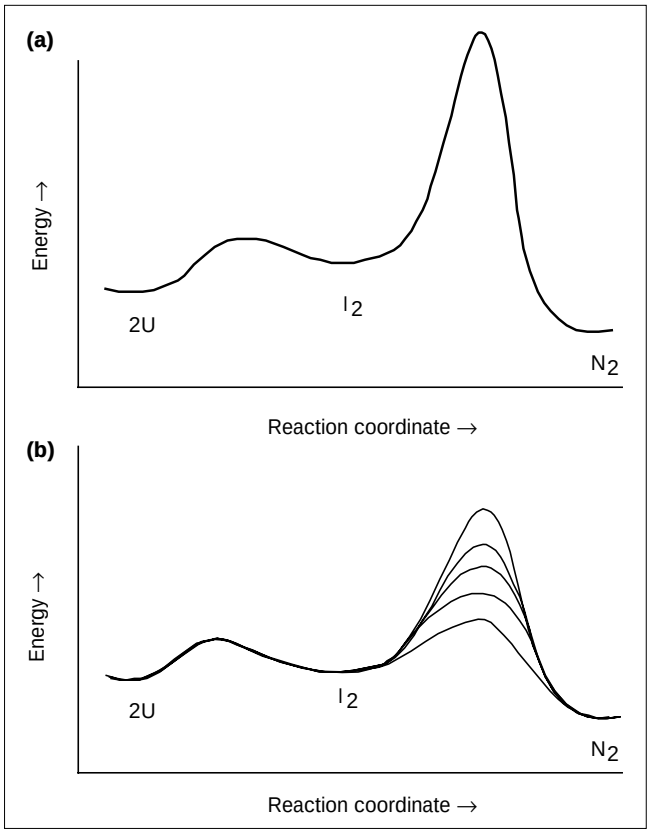
effects of the repacked layers are additive. As the number of repacked layers increases, the folding and unfolding rates increase in parallel (Fig. 10b). Third, superimposed upon effects 1 and 2 is an 'end effect'. Replacing the wild-type-like 'reversal' in layers 2 and 7 (Leu_2Ala_2) with Ala_2Leu_2 increases the rates of folding and unfolding approximately 10-fold. This effect holds regardless of context (Fig. 10c).

The rate enhancements primarily reflect the effects of the hydrophobic core mutations on the energy of the transition state. Differences in the stabilities of the native state and intermediate will, of course, also influence the rates of unfolding and folding, respectively. Although we do not know to what extent the mutations change these energies, we do know that the mutations cause relatively small changes in the difference in free energy between the folded and unfolded states. The effects on folding and unfolding rates of such stability differences are superimposed upon transition state effects, such that the above three 'rules' do not precisely quantitatively describe the effect of every mutant.

What is the molecular basis of the rate enhancements? Repacking any single layer causes a rate enhancement. This suggests that either alanine and leucine pack more readily within a layer, or that alanine-leucine layers pack more readily with neighboring layers. The residues in wild type that are not alanine or leucine are Glu5, Ile15, Thr19, Gln34, Cys38, Cys52, and Phe56. The crystal structure of wild-type Rop reveals that the polar residues pack with the hydrophilic portions of their sidechains outside the core with only their hydrophobic portions buried [26]. In addition, several of these residues are involved in specific hydrogen-bonding interactions. Sidechain interactions include Glu5 with Thr2, Thr19 with His42, Gln34 with Ser51, and Arg55 with both Gln34 and Asp32. In addition, the mainchain carbonyl of both Ile15 and Arg16 interact with the sidechain of Thr19, and the mainchain amide of Glu5 with the sidechain of Thr2. These polar interactions and the particular packing associated with them could take much longer to achieve, and be more difficult to unlock, than alanine-leucine packing. Replacement of partially buried salt bridges and hydrogen-bonding interactions has recently been shown to increase the rate of folding of the Arc repressor protein [14]. For the hydrophobic Ile15, Cys38, Cys52, and Phe56 sidechains, it is possible that they take longer to align correctly since they have less freedom to explore sidechain rotamer conformations in an α-helix than do alanine and leucine.

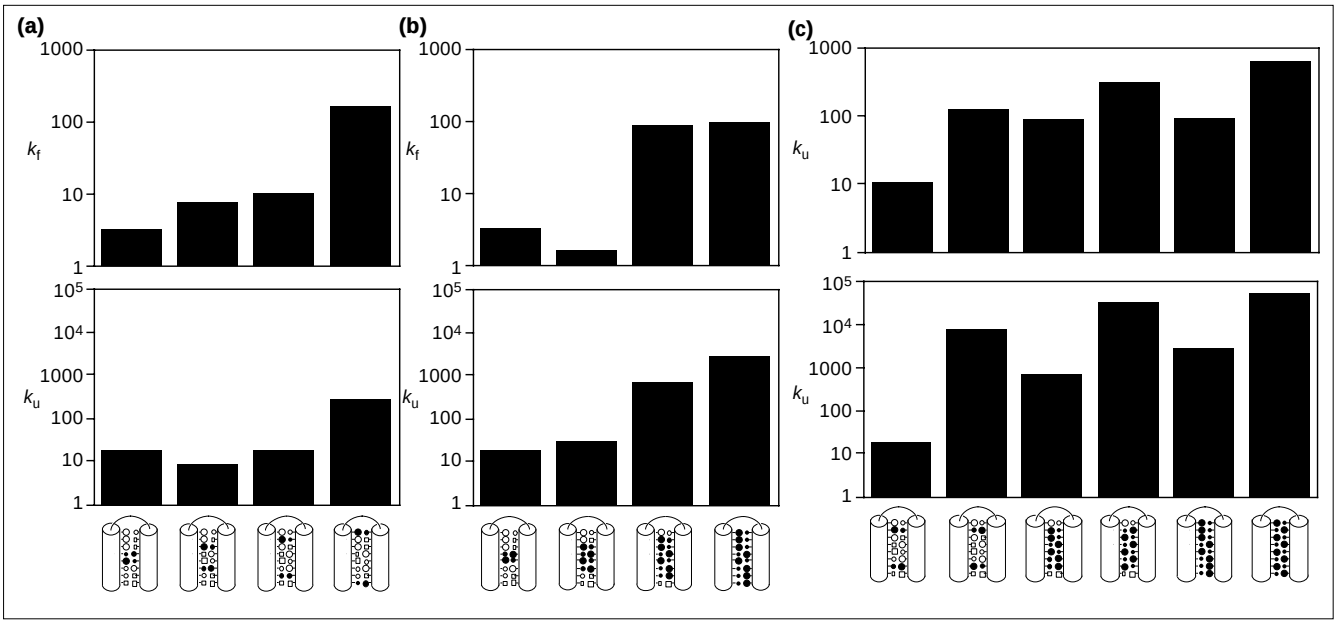
Finally, removing the 'reversed' layer 2 and 7 packing causes an additional enhancement in the folding and unfolding rate. The reversed configuration itself may be more difficult to align correctly or to pack against the somewhat unusual first and eighth layer, in which Phe56 is actually coming in from an 'e' position. Conversely, in unfolding, it may be relatively more difficult to disrupt this unusual end layer packing and for denaturant to enter the core than it is to disrupt an Ala_2Leu_2 layer.

Figure 9



Schematic illustration of a possible reaction coordinate for the folding of (a) wild-type Rop and (b) the repacked mutants. For wild-type Rop, the activation barrier between 2U and I_2 is relatively low – this step occurs rapidly. The activation barrier between I_2 and N_2 is high – this step occurs slowly and is rate-limiting for both folding and unfolding. The main effect of the hydrophobic core mutants is to decrease, to varying degrees, the activation barrier of the I_2 to N_2 transition, as illustrated. These diagrams are intended to illustrate the main features of our model and are not comprehensive in detail. Omitted for clarity are the following details: the equilibrium stabilities of the wild-type and the repacked proteins are not identical, and the stabilites of the unfolded state and the intermediate may also differ, but we have drawn them to be the same. The rate of the fast phase is increased for some, perhaps all, of the mutants. We were unable to quantify this effect and have therefore drawn the activation barrier as being lower for the mutants than the wild type, but have not distinguished between the different mutants. We have drawn the position on the reaction coordinate of I_2 for the mutants to be the same as for the wild type. Our data indicate that I_2 for the mutants is slightly different for the mutants than for the wild type, with approximately 60% burial of hydrophobic residues relative to the native structure in the wild type and 70–75% in the mutants. The estimates of the percentage burial of hydrophobic surface are based on the ratios of m_f/m, the slopes of plots of $\ln k_f$ versus [GdnHCl] and $\ln K_{eq}$ versus [GdnHCl], respectively.

Figure 10



Bar graphs comparing the rates of folding and unfolding of the different repacked mutants relative to the wild-type rate. (a) Mutants with one layer changed per monomer at different positions within the core. (b) Mutants with increasing numbers of repacked layers, with the wild-type-like layer 2 and 7 reversal retained. (c) Rates for mutants repacked with either Leu_2Ala_2 or Ala_2Leu_2 in layers 2 and 7.

Summary

Wild-type Rop displays unusually slow folding and unfolding kinetics. The studies we report identify the nature of the rate-limiting step and demonstrate quantitatively how repacking the hydrophobic core lowers this activation barrier, thereby increasing both folding and unfolding rates. Our results show dramatic increases in rates of up to two orders of magnitude for folding and five orders of magnitude for unfolding. The magnitude of these enhancements is unprecedented. Although the slow kinetics exhibited by wild-type Rop were certainly not anticipated, our results allow us to begin to rationalize possible causes. Studies are in progress to further delineate the slow folding pathway and to address its significance *in vivo*.

Materials and methods

Cloning, protein purification and characterization

The genes encoding wild-type Rop and the repacked mutants were created using a combination of chemical oligonucleotide synthesis and PCR and cloned into the T7 expression vector pMR101 [27]. The sequences of the proteins (using the one-letter amino acid code) are as follows: wild-type Rop (GTKQEKTALNMARFIRSQTLTLLEKLNELDAD-EQADICESLHDHADELYRSCLARFGDDGENL); Ala_2Leu_2-2: I15L, T19A; Ala_2Leu_2-(3+6): C38A; Leu_2Ala_2-(2+7): Q34A, C52L; Ala_2Leu_2-(1+8): E5A, R55L, F56A; Ala_2Leu_2-4: I15L, T19A, C38A; Ala_2Leu_2-6-rev: I15L, T19A, Q34A, C38A, C52L; Ala_2Leu_2-8-rev: E5A, I15L, T19A, Q34A, C38A, C52L, R55L, F56A; Ala_2Leu_2-(2+7): A8L, L26A, Q34L, C52A; Ala_2Leu_2-6: I15L, T19A, Q34L, C38A, C52A; Ala_2Leu_2-8: E5A, I15L, T19A, Q34L, C38A, C52A, R55L, F56A. Wild-type Rop and each of the repacked mutants were overproduced, purified, and tested for retention of RNA-binding activity as previously described [18,24].

CD measurements

Ellipticity at 222 nm was measured as a function of temperature or concentration of GdnHCl (Pierce) using an AVIV Model 62DS CD Spectropolarimeter (AVIV Instruments, Lakewood, NJ). The concentration of each protein was 20 μM dimer in 100 mM sodium phosphate, pH 7, and 200 mM NaCl (plus 1 mM DTT if the protein contained cysteine residues). Thermal and GdnHCl denaturations were performed and analyzed as described previously [18,19]. The thermal and chemical denaturation transitions were reversible for all the proteins.

For the wild-type protein and the slowly unfolding mutants, unfolding reactions were monitored by CD at 25°C using the buffer and protein concentration described above. Folded protein in phosphate buffer was quickly (<30 s) added to strongly unfolding concentrations of GdnHCl, and the ellipticity at 222 nm recorded at appropriate intervals during the unfolding process. To allow a meaningful comparison, because the mutant and wild-type proteins all have slightly different stabilities, we report the rate of folding at the concentration of GdnHCl at which the protein would be 90% folded and the rate of unfolding at the concentration of GdnHCl at which the protein would be 90% unfolded. These concentrations were chosen so that the reverse reaction does not contribute significantly.

Curve fitting was performed using the program KaleidaGraph (distributed by Synergy Software); the unfolding curves were well fit by a single exponential, as determined by visual inspection and low χ^2 values:

$$Y = A_1 e^{-k_1 t} + C \quad (1)$$

where Y is the fluorescence intensity at time t, A_1 is the amplitude change during the unfolding reaction, and k_1 is the unfolding rate.

The effect of temperature on the unfolding of wild type in 5.0 M GdnHCl and the Ala_2Leu_2-4 protein in 4.0 M GdnHCl was analyzed using the Arrhenius equation [28]:

$$k = A e^{-(E_a/RT)} \quad (2)$$

When 1/T was plotted against ln*k*, and the data linearly fit, the values for E_a were calculated from the slope of the line (E_a/R).

Stopped-flow fluorescence measurements

For the fast unfolding mutants, and for all the refolding experiments, stopped-flow measurements were performed using a Kintek SF-2001 (Kintek Instruments, State College, PA) fluorimeter thermostatted at 20°C with a 0.5 cm pathlength. 5 ml and 0.5 ml syringes were used to mix the buffer/GdnHCl solutions and protein solutions at a 10:1 ratio. The buffer was the same as for the CD measurements, but the final protein concentration used was 100 μM. For the refolding experiments, slowly unfolding proteins such as the wild type were incubated 12–24 h in high concentrations of GdnHCl to insure their complete denaturation. Intrinsic tyrosine fluorescence was excited at 275 nm and monitored with a filter cut off of greater than 289 nm. The data for 2–6 runs were averaged for analysis using software provided by Kintek Instruments; more runs were needed to improve the signal : noise ratios for the faster unfolding/folding mutants. The unfolding data for Ala_2Leu_2-2+7, Ala_2Leu_2-6, and Ala_2Leu_2-8 fit well to a single exponential (eq. 1). Refolding data were analyzed using either a single exponential or a double exponential:

$$Y = A_1 e^{-k_1 t} + A_2 e^{-k_2 t} + C \quad (3)$$

Stopped-flow CD measurements

Measurements were made on a Biologic stopped-flow spectropolarimeter. Unfolded protein was prepared in 100 mM Na phosphate, pH 7, 200 mM NaCl, 1 mM DTT, 6.4 M GdnHCl. Refolding was initiated by diluting the protein 10-fold into the same buffer lacking GdnHCl. The starting concentration of denatured protein was approximately 700 μM. Multiple repetitions of the folding reaction were performed (of the order 40–100) and the results averaged to improve the signal : noise ratio. Under the conditions of these experiments, mixing artefacts prevented the acquisition of meaningful data until approximately 1 s after folding was initiated.

Acknowledgements

We thank Phillip Evans, Department of Biochemistry, University of Cambridge, UK, for discussions and for generously allowing us to use his stopped-flow CD. We thank Julian Sturtevant for his continued interest and for his DSC studies of many Rop variants. We thank Athena Nagi, Jenny Yang, Michael Mahoney and Elizabeth Doherty for sharing unpublished data, and we thank them and other members of the Regan group for critical reading of the manuscript and discussions. This work was supported by grant GM49146. L Regan is an NSF National Young Investigator and a Dreyfus Teacher-Scholar.

References

1. Levinthal, C. (1968). Are there pathways for protein folding? *J. Chim. Phys.* 65, 44–45.
2. Baldwin, R. (1996). Why is protein folding so fast? *Proc. Natl. Acad. Sci. USA* 93, 2627–2628.
3. Beasty, A., Hurle, M., Manz, J., Stackhouse, T., Onuffer, J. & Matthews, C. (1986). Effects of the phenylalanine-22 → leucine, glutamic acid-49 → methionine, glycine-234 → aspartic acid, and glycine-234 → lysine mutations on the folding and stability of the α subunit of tryptophan synthase from *Escherichia coli*. *Biochemistry* 25, 2965–2974.
4. Elöve, G.A., Chaffotte, A.F., Roder, H. & Goldberg, M.E. (1992). Early steps in cytochrome *c* folding probed by time-resolved circular dichroism and fluorescence spectroscopy. *Biochemistry* 31, 6876–6883.
5. Kim, P. & Baldwin, R. (1990). Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* 59, 631–660.
6. Jackson, S.E. & Fersht, A.R. (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* 30, 10428–10435.
7. Chen, B.-L., Baase, W.A., Nicholson, H. & Schellman, J.A. (1992). Folding kinetics of T4 lysozyme and nine mutants at 12°C. *Biochemistry* 31, 1464–1476.
8. Serrano, L., Matouschek, A. & Fersht, A. (1992). The folding of an enzyme III. Structure of the transition state for unfolding of barnase analyzed by a protein engineering procedure. *J. Mol. Biol.* 224, 805–818.
9. Schindler, T., Herrler, M., Marahiel, M. & Schmid, F. (1995). Extremely rapid protein folding in the absence of intermediates. *Nat. Struct. Biol.* 2, 663–673.
10. Gittelman, M.S. & Matthews, C.R. (1990). Folding and stability of *trp* aporepressor from *Escherichia coli*. *Biochemistry* 29, 7011–7020.
11. Milla, M.E. & Sauer, R.T. (1994). P22 Arc repressor: folding kinetics of a single-domain, dimeric protein. *Biochemistry* 33, 1125–1133.
12. Wendt, H., Berger, C., Baici, A., Thomas, R.M. & Bosshard, H.R. (1995). Kinetics of folding of leucine zipper domains. *Biochemistry* 34, 4097–4107.
13. Zitzewitz, J.A., Bilsel, O., Luo, J., Jones, B.E. & Matthews, C.R. (1995). Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. *Biochemistry* 34, 12812–12819.
14. Waldburger, C., Jonsson, T. & Sauer, R. (1996). Barriers to protein folding: formation of buried polar interactions is a slow step in acquisition of structure. *Proc. Natl. Acad. Sci. USA* 93, 2629–2634.
15. Mok, Y.-K., Bycroft, M. & De Prat Gay, G. (1996). The dimeric DNA binding domain of the human papillomavirus E2 protein folds through a monomeric intermediate which cannot be native-like. *Nat. Struct. Biol.* 3, 711–717.
16. Cesareni, G. & Banner, D.W. (1985). Regulation of plasmid copy number by complementary RNAs. *Trends Biochem. Sci.* 17, 303–306.
17. Marino, J.P., Gregorian, R.S., Jr., Csankovszki, G. & Crothers, D.M. (1995). Bent helix formation between RNA hairpins with complementary loops. *Science* 268, 1448–1454.
18. Munson, M., O'Brien, R., Sturtevant, J.M. & Regan, L. (1994). Redesigning the hydrophobic core of a four-helix-bundle protein. *Protein Sci.* 3, 2015–2022.
19. Munson, M., *et al*., & Regan, L. (1996). What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. *Protein Sci.* 5, 1584–1593.
20. Segawa, S.-I., Husimi, Y. & Wada, A. (1973). Kinetics of thermal unfolding of lysozyme. *Biopolymers* 12, 2521–2537.
21. Dobson, C. & Evans, P. (1984). Protein folding kinetics from magnetization transfer nuclear magnetic resonance. *Biochemistry* 23, 4267–4270.
22. Milla, M.E., Brown, B.M., Waldburger, C.D. & Sauer, R.T. (1995). P22 Arc pepressor: transition state properties inferred from mutational effects on the rates of protein unfolding and refolding. *Biochemistry* 34, 13914–13919.
23. Chen, B.-L., Baase, W. & Schellman, J. (1989). Low-temperature unfolding of a mutant of phage T4 lysozyme. 2. Kinetic investigations. *Biochemistry* 28, 691–699.
24. Predki, P.F., Nayak, L.M., Gottlieb, M.B. & Regan, L. (1995). Dissecting RNA–protein interactions: RNA–RNA recognition by Rop. *Cell* 80, 41–50.
25. Burns, J., Butler, J., Moran, J. & Whitesides, G. (1991). Selective reduction of disulfides by tris(2-carboxyethyl) phosphine. *J. Org. Chem.* 56, 2648–2650.
26. Banner, D., Kokkinidis, M. & Tsernoglou, D. (1987). Structure of the ColE1 Rop protein at 1.7 Å resolution. *J. Mol. Biol.* 196, 657–675.
27. Munson, M., Predki, P.F. & Regan, L. (1994). ColE1-compatible vectors for high-level expression of cloned DNAs from the T7 promoter. *Gene* 144, 59–62.
28. Tinoco, I., Jr, Sauer, K. & Wang, J. (1985). *Physical Chemistry: Principles and Applications in Biological Sciences*. Englewood Cliffs, NJ, Prentice-Hall, Inc.
29. Kraulis, P.J. (1991). Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946–950.

Because *Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published via the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/fad.htm – for further information, see the explanation on the contents page.