Role of a solvent-exposed aromatic cluster in the folding of *Escherichia coli* CspA

HECTOR M. RODRIGUEZ, DUNG M. VU, AND LYDIA M. GREGORET

Department of Chemistry and Biochemistry, University of California, Santa Cruz, California 95064

(RECEIVED March 20, 2000; FINAL REVISION July 13, 2000; ACCEPTED July 28, 2000)

Abstract

Escherichia coli CspA is a member of the cold shock protein family. All cold shock proteins studied to date fold rapidly by an apparent two-state mechanism. CspA contains an unusual cluster of aromatic amino acids on its surface that is necessary for nucleic acid binding and also provides stability to CspA (Hillier et al., 1998). To elucidate the role this aromatic cluster plays in the determining the folding rate and pathway of CspA, we have studied the folding kinetics of mutants containing either leucine or serine substituted for Phe18, Phe20, and/or Phe31. The leucine substitutions are found to accelerate folding and the serine substitutions to decelerate folding. Because these residues exert effects on the free energy of the folding transition state, they may be necessary for nucleating folding. They are not responsible, however, for the very compact, native-like transition state ensemble seen in the cold shock proteins, as the refolding rates of the mutants all show a similar, weak dependence of unfolding rate on denaturant concentration. Using mutant cycle analysis, we show that there is energetic coupling among the three residues between the unfolded and transition states, suggesting that the cooperative nature of these interactions helps to determine the unfolding rate. Overall, our results suggest that separate evolutionary pressures can act simultaneously on the same group of residues to maintain function, stability, and folding rate.

Keywords: aromatic side-chain interactions; β -sheet formation; phi-value analysis; protein folding kinetics; rapid folding; thermodynamic mutant cycle

Understanding the mechanisms by which proteins achieve their three-dimensional (3D) structures is a fundamental problem that has captured the interest of protein chemists, molecular biologists, and theoreticians for the last half of the twentieth century. The recent discovery that small, single-domain proteins can fold rapidly and by an apparent two-state mechanism (reviewed in Jackson, 1998) has led to studies and discussion about whether there are critical intrachain interactions that are prerequisites for productive folding (Fersht et al., 1992; Abkevich et al., 1994; Itzhaki et al., 1995; Grantcharova et al., 1998; Martinez et al., 1998a). There is intense interest in these specific, nucleating interactions because their identification could potentially reveal the elusive connection between a protein's primary sequence and its 3D structure.

The cold shock proteins are paradigms of simple, two-state folding reactions (Schindler et al., 1995, 1999; Perl et al., 1998; Reid et al., 1998). Cold shock proteins fold very rapidly, with time constants of 5 ms or less. We have shown previously that three clustered phenylalanine residues that are part of the nucleic acid binding site of *Escherichia coli* CspA (Fig. 1) contribute signifi-

cantly to this protein's stability and that there are cooperative, stabilizing interactions between these residues (Hillier et al., 1998).

Aromatic:aromatic interactions are commonly observed on adjacent strands of antiparallel β -sheets (Wouters & Curmi, 1995; Hutchinson et al., 1998) and have been shown to contribute to protein stability (Smith & Regan, 1995; Hillier et al., 1998). While to our knowledge, no kinetic studies have (until now) been performed to probe the importance of this type of interaction for β -sheet protein folding, it has been speculated that such interactions may help guide β -strand pairing and therefore be important for efficient folding (Tisi & Evans, 1995; Hillier et al., 1998).

Here, we report on the folding kinetics of mutants of CspA in which neighboring phenylalanine residues in this cluster have been mutated to serine and to leucine. These mutations were found to affect both the rates of protein refolding and unfolding, suggesting that these residues play a role in determining the fast folding behavior of CspA. Interestingly, we find that in spite of the fact that they reduce protein stability, two of the Phe $\rightarrow$ Leu substitutions actually accelerate both refolding and unfolding. Combining these two leucine substitutions leads to further acceleration. In the case of the serine mutants, the destabilization is due primarily to a reduction in the refolding rates, suggesting a shift in the free energy of the transition state. We have also now extended our previous study on the cooperativity between these residues (Hillier

Reprint requests to: Lydia M. Gregoret, Department of Chemistry and Biochemistry, University of California, Santa Cruz, California 95064; e-mail: gregoret@chemistry.ucsc.edu.

Fig. 1. Ribbon diagram of *E. coli* CspA showing the aromatic cluster with the phenylalanine residues highlighted. This image was rendered with the Ribbons software (Carson, 1991) using the coordinates of Schindelin et al. (1994). Individual β -strands are numbered in bold.

et al., 1998) by performing mutant cycle analysis using the data obtained from folding kinetics experiments.

Although the solvent-exposed aromatic residues in CspA may account, in part, for the fast folding rates seen for the cold shock proteins, mutation of these residues does not cause major changes in the solvent accessibility of the transition state ensemble. Thus, the solvent-exposed aromatic cluster is most likely *not* responsible for the unusually compact transition state structure that is characteristic of the cold shock proteins (Schindler et al., 1995, 1999; Perl et al., 1998; Reid et al., 1998). Based on our results, we propose several mechanisms for the assembly of structure in CspA and discuss their implications to β -sheet protein folding and protein evolution.

Results

Residues F18, F20, and F31 of *E. coli* CspA were mutated both singly and in combination to leucine and to serine as described previously by Hillier et al. (1998). The mutants are designated by a three-letter code referring to the amino acid residues at positions 18, 20, and 31, respectively. For example, mutant "FLF" refers to a single mutant with Phe20 substituted with leucine. Mutant proteins containing alanine at these three positions were also created but, unfortunately, could not be produced in sufficient quantities for the kinetics studies.

The folding kinetics of all possible leucine mutants and of all serine single mutants were investigated by stopped flow fluorescence spectroscopy using urea as a denaturant. Refolding was

initiated by rapid dilution of concentrated protein in 7 M urea into refolding buffer. Unfolding was initiated by mixing concentrated protein in buffer into buffered urea. We chose urea for these experiments instead of guanidinium chloride (GdmCl) because in wild-type CspA, the denaturation midpoint occurs at a higher concentration for urea (4.0 M) than for GdmCl (1.5 M) (Reid et al., 1998). This affords us a greater concentration range within which to observe refolding-dominated reactions, especially for destabilized mutants.

In all cases, the fluorescence signal is fit well by a monoexponential time course suggesting a two-state folding mechanism, consistent with the behavior of wild-type CspA (Reid et al., 1998). To obtain the refolding and unfolding rate constants in the absence of denaturant, the folding experiments were performed at a variety of final urea concentrations. The apparent rate constants (λ) derived from monoexponential fits to the time course of the fluorescence signal were plotted as a function of urea concentration. Figure 2 shows the dependence of the folding rate on urea concentration for the leucine single mutants (A), the leucine multiple mutants (B), and the serine single mutants (C). The values of the folding rates and denaturant dependencies are shown in Table 1. Within the limits of our instrument, the full signal amplitude of the slower folding proteins was recovered. However, as the folding rates increased, it became difficult to interpret amplitude results.

When residue 18 is substituted with leucine, the resulting protein unfolds more rapidly than the wild-type protein, but refolds with a comparable rate. Substitutions F20L ("FLF") and F31L ("FLL") cause 1.5–2-fold increases in the refolding rate of CspA (Fig. 2A). The leucine substitutions also cause 3–5.5-fold increases in the unfolding rate of CspA.

When leucine substitutions at positions 20 and 31 are combined in mutant FLL, the folding rates increase further (608 s⁻¹ for refolding and 13 s⁻¹ for unfolding; Fig. 2B). The values of the folding rates for FLL and for the other multiple leucine mutants should be interpreted with caution, however. The measured apparent rate constants of these mutants are above 100 s⁻¹ at many urea concentrations (and at *all* concentrations in the case of LLL). We estimate that the dead time of our instrument is two milliseconds. Thus, apparent reaction rates faster than about 350 s⁻¹ (the respective rate constant) cannot be measured accurately, and rates approaching that value will have a greater error associated with them.

The refolding rates of the three serine single mutants in the absence of urea are twofold to fourfold slower than that of wild-type CspA (Fig. 2C; Table 1). The refolding rate is slowest in mutant FFS. The unfolding rates of these three mutants are closer to wild-type CspA than those of the leucine mutants.

One unusual characteristic of the folding of all cold shock proteins studied thus far is a very slight dependence of the unfolding rate on urea concentration (Schindler et al., 1995, 1999; Perl et al., 1998; Reid et al., 1998). In other words, the unfolding (right) branches of the apparent rate vs. urea plots (chevron plots) have very modest slopes. In general, a strong dependence on denaturant concentration is thought to reflect a large difference in the solvent accessibility of two ensembles (e.g., the folded and unfolded ensembles) (Schellman, 1978). The slight dependence of the unfolding rate on urea concentration implies that the transition and folded state ensembles of the cold shock proteins have similar solvent accessibilities. A quantitative estimate of the relative difference in solvent accessibilities of the transition and folded states of a protein may be inferred from the value of $\alpha^\ddagger$ (Equation 2; Table 2).

Fig. 2. Refolding rates. **A:** Leucine single mutants. **B:** Leucine multiple mutants. **C:** Serine single mutants as a function of urea concentration determined by stopped flow fluorescence spectroscopy at 25 °C. Curve fits were obtained as described in Materials and methods.

This is the fraction of the overall denaturant dependence of the unfolding transition that is accounted for between the unfolded state and the transition state (see Materials and methods). By this measure, the transition state of wild-type CspA is 86% native-like (or, 14% unfolded). The leucine and serine mutants, in general, also have very native-like transition states. Although it appears as though the mutants under study here have more symmetrical,

V-shaped urea dependences, this is an illusion that arises from the fact that more of the unfolding arm can be seen in these destabilized mutants. The only mutant for which $\alpha^\ddagger$ is significantly smaller is FLL, which is also the fastest refolding mutant. The solvent accessibility of the transition state ensemble of FLL appears to be shifted slightly toward the unfolded state (69% native-like).

The relative free energies of the folding transition states of the CspA mutants can be computed using ϕ -value analysis (Fersht, 1995). This method is the principal means of identifying amino acid residues that form native-like, stabilizing interactions in the folding transition state using mutagenesis. The ϕ -value is the ratio of free energy difference of the transition states ($\ddagger$) relative to the overall free energy difference between a mutant and wild-type protein (Equation 3). We adopt the convention that the ϕ -value is computed in the refolding direction (i.e., $\phi_f = \Delta\Delta G_{\ddagger-U} / \Delta\Delta G_{F-U}$). A residue that stabilizes the folded form and the transition state to the same extent will have a ϕ_f -value of 1.0. Such a residue is thought to make native-like interactions in the transition state and is likely to be ordered in the rate-limiting step of folding. A residue that forms stabilizing interactions *after* passing through the transition state will have a ϕ_f -value of 0. Such residues do not participate in the rate-limiting step. Fractional ϕ_f -values are relatively common and may arise from either partially formed, native-like interactions, or multiple, parallel folding pathways. Negative ϕ_f -values are very rarely observed and may be indicative of nonnative, stabilizing interactions in the transition state (Martinez et al., 1998a).

Although ϕ_f -values are often computed for mutations to alanine, given the large size difference between phenylalanine and alanine, such a substitution could potentially lead to structural rearrangements that result in different side-chain-side-chain interactions in the mutant that do not exist in the wild-type protein. A phenylalanine to leucine substitution, however, removes only an aromatic ring, but retains the C_γ and C_δ carbons, and is thus more conservative.

The ϕ -values for the leucine single substitutions are shown in Table 2. Only the unfolding rate is affected in mutant LFF. Hence, this mutation has little effect on the value of ϕ_f . Mutants FLF and FFL, however, refold faster than wild-type CspA. Their unfolding rates are also faster. The effects of these changes result in significant negative ϕ -values of -0.57 and -0.69 , respectively (referring to ϕ_f values computed using ΔG obtained from kinetic analysis). The latter value, however, may not be as meaningful since the free energy change upon mutation of Phe31 to leucine is only 0.3 kcal/mol. The ϕ_f -values calculated for the corresponding serine mutants are positive and in the case of FFS are quite large ($\phi_f = +0.81$). However, because the serine side chain has a hydroxyl group, there is the possibility of introducing a new type of interaction (e.g., a hydrogen bond) not present in the wild-type protein.

To examine whether there are cooperative interactions between the phenylalanine residues of CspA, we used the kinetic data on the folding rates of the leucine mutants of CspA to compute thermodynamic mutant cycles. The results of these calculations are shown in Figure 3. The energetic coupling calculated using the free energies obtained by kinetic analysis are labeled $^B\Delta G_{int}$. These coupling energies are similar to those previously obtained in the equilibrium analysis of these residues ($^A\Delta G_{int}$) (Hillier et al., 1998), and reinforce our earlier conjecture that the residues of aromatic cluster on the surface of CspA interact cooperatively to stabilize the protein. Although the interaction energy between residues F18 and F20 appears to be smaller than obtained previously (0.1 vs. 0.4 kcal/mol), it is nonetheless within error of the earlier value. (The

Table 1. Thermodynamic and kinetic parameters of wild-type CspA (FFF) and mutants

	T_m (°C)	ΔG_{eq}^a (kcal mol ⁻¹)	ΔG_{kin} (kcal mol ⁻¹)	k_f (s ⁻¹)	m_f (M ⁻¹)	k_u (s ⁻¹)	m_u (M ⁻¹)	m_{eq}^a (kcal mol ⁻¹ M ⁻¹)	m_{kin} (kcal mol ⁻¹ M ⁻¹)
FFF	58 ± <1	3.0 ± 0.2	2.9 ± 0.1	264 ± 16	-0.87 ± 0.02	1.9 ± 0.2	0.14 ± 0.07	0.71 ± 0.04	0.60 ± 0.07
LFF	48 ± <1	2.0 ± 0.1	2.0 ± 0.1	247 ± 8	-1.00 ± 0.04	8.4 ± 0.8	0.23 ± 0.02	0.80 ± 0.04	0.73 ± 0.04
FLF	58 ± <1	2.7 ± 0.2	2.5 ± 0.2	379 ± 25	-0.71 ± 0.03	5.9 ± 1.7	0.30 ± 0.01	0.71 ± 0.05	0.60 ± 0.03
FFL	56 ± <1	2.3 ± 0.1	2.3 ± 0.1	499 ± 33	-0.90 ± 0.03	11.1 ± 0.6	0.22 ± 0.01	0.72 ± 0.04	0.66 ± 0.03
FLL	59 ± <1	2.5 ± 0.1	2.3 ± 0.2	608 ± 46	-0.64 ± 0.04	13.0 ± 3.4	0.29 ± 0.05	0.70 ± 0.01	0.55 ± 0.07
LFL	48 ± <1	1.4 ± 0.1	1.5 ± 0.1	340 ± 28	-0.91 ± 0.06	29.5 ± 2.3	0.20 ± 0.01	0.72 ± 0.01	0.66 ± 0.05
LLF	55 ± <1	2.1 ± <0.1	1.6 ± 0.5	381 ± 231	-0.82 ± 0.90	24.9 ± 12.0	0.25 ± 0.05	0.72 ± 0.01	0.63 ± 0.84
LLL	58 ± <1	2.1 ± 0.1	1.7 ± 0.9	344 ± 164	-0.36 ± 0.23	21.3 ± 31.3	0.23 ± 0.13	0.66 ± 0.01	0.35 ± 0.45
SFF	45 ± <1	1.5 ± <0.1	1.9 ± 0.1	91 ± 7	-0.86 ± 0.06	3.5 ± 0.7	0.36 ± 0.09	0.76 ± 0.01	0.72 ± 0.09
FSF	48 ± <1	1.8 ± <0.1	1.8 ± <0.1	129 ± 5	-1.25 ± 0.04	6.8 ± 0.3	0.24 ± 0.11	0.78 ± 0.01	0.88 ± 0.08
FFS	48 ± <1	2.0 ± <0.1	2.0 ± <0.1	65 ± 1	-1.12 ± 0.02	2.2 ± 0.1	0.24 ± 0.02	0.80 ± 0.01	0.81 ± 0.02

Note: Mutant abbreviations are explained in the text.

^aValues taken from Hillier et al. (1998).

other face of the “mutant cube” is not shown because of the large errors associated with the folding rates of mutants LLL and LLF. However, the interaction energies computed for these mutants are consistent with those for the single and double mutants.)

To determine whether the energetic coupling between the phenylalanines of the aromatic cluster derives from the unfolding or refolding portion of the reaction, we computed the free energy differences between the unfolded state and the transition state and between the folded state and transition state using the rates determined by the stopped flow experiments (Equations 7 and 8). As can be seen in Figure 3, the interaction energies between F20 and F31 and F18 and F31 are significant for the unfolding reaction ($^D\Delta G_{int}^\ddagger$). Removing one member of an interacting pair (e.g., F20 of the F20–F31 pair) causes a significant increase in the unfolding rate. However, subsequently mutating the other member of the pair has a lesser effect on the rate, regardless of which mutation is made first. This indicates that the two residues interact in a cooperative fashion to determine the unfolding rate of CspA. For the cycle relating to the

refolding transition state, none of the residues under consideration appear to interact cooperatively to determine the refolding rate, since all of the interaction energies are close to zero ($^C\Delta G_{int}^\ddagger$).

We had previously characterized the stability of these mutants using urea denaturation (Hillier et al., 1998). We have now also investigated their stability to thermal denaturation by monitoring the change in molar ellipticity as a function of temperature by circular dichroism (CD) spectroscopy (Table 1). For some destabilized mutants, the midpoint of the thermal denaturation transition (T_m) is the same as wild-type CspA. For example, the triple mutant LLL has an unfolding free energy of 2.1 kcal/mol, 0.9 kcal/mol lower than wild-type CspA when measured by urea denaturation. However, this mutant has an identical T_m to wild-type CspA, suggesting that either the curvature or the maximum of the free energy vs. temperature profile for LLL and wild-type CspA is different.

Discussion

The role of the aromatic cluster in the stability and folding of CspA

We had previously shown that three of the aromatic cluster residues of CspA—F18, F20, and F31—are important for nucleic acid

Table 2. Transition state characteristics of wild-type CspA and its mutants

	$\alpha_{kin}^\ddagger$ ^a	$\alpha_{eq}^\ddagger$ ^b	$\phi_{f(kin)}^\ddagger$ ^c	$\phi_{f(eq)}^\ddagger$ ^d
FFF	0.86 ± 0.07	0.73 ± 0.04	—	—
LFF	0.81 ± 0.04	0.74 ± 0.05	0.04 ± 0.04	0.04 ± 0.04
FLF	0.75 ± 0.04	0.59 ± 0.05	-0.69 ± 0.52	-0.47 ± 0.22
FFL	0.80 ± 0.04	0.74 ± 0.05	-0.57 ± 0.18	-0.57 ± 0.10
FLL	0.69 ± 0.06	0.54 ± 0.03	-1.02 ± 0.36	-0.77 ± 0.22
LFL	0.82 ± 0.07	0.75 ± 0.05	-0.10 ± 0.04	-0.10 ± 0.04
LLF	0.77 ± 1.06	0.67 ± 0.74	-0.25 ± 0.41	-0.17 ± 0.28
LLL	0.61 ± 0.48	0.32 ± 0.21	-0.18 ± 0.33	-0.12 ± 0.24
SFF	0.70 ± 0.08	0.67 ± 0.05	0.41 ± 0.06	0.64 ± 0.11
FSF	0.84 ± 0.07	0.95 ± 0.03	0.37 ± 0.06	0.36 ± 0.04
FFS	0.82 ± 0.02	0.83 ± 0.02	0.81 ± 0.12	0.92 ± 0.08

^a $\alpha_{kin}^\ddagger$, given by Equation 2 using m_u and m_f values for the respective protein.

^b $\alpha_{eq}^\ddagger$, given by Equation 2 using m_{eq} value from Hillier et al. (1998).

^c $\phi_{f(kin)}^\ddagger$, given by Equation 3 using ΔG_U computed from Equation 6.

^d $\phi_{f(eq)}^\ddagger$, given by Equation 3 using ΔG_U values from Hillier et al. (1998).

Fig. 3. Double mutant cycles of residues F18, F20, and F31. **A:** Cycle from equilibrium data reported in Hillier et al. (1998). **B:** Cycle computed from free energy of unfolding determined from the kinetic data. **C:** Cycle for the free energy between the unfolded state and transition state. **D:** Cycle for the free energy difference between the folded state and the transition state. The interaction energies (ΔG_{int} kcal/mol) are calculated as differences in the free energy between parallel arrows (see Materials and methods).

binding and that they each contribute 0.3–1.0 kcal/mol to the stability of the protein in comparison with leucine at these positions (Hillier et al., 1998). Substitution with serine is more severely destabilizing ($\Delta\Delta G = 1.0$ – 1.5 kcal/mol). We also found cooperative, stabilizing interactions between F18 and F20 and between F20 and F31. We have now studied the effects of these mutations on the folding kinetics of CspA.

We find that the serine substitutions, which are most detrimental to stability and binding, also decelerate refolding of CspA twofold to fourfold (Fig. 2C; Table 1). This leads to ϕ -values greater than 0 and, in the case of FFS, approaching 1.0 (Table 2). There are two possible explanations for the slowed refolding rates observed in these mutants. One is that the unfolded states of the serine mutants are stabilized relative to the wild-type protein (but that the transition state energies are unchanged). It seems unlikely that single substitutions such as these, though nonconservative, would greatly affect the free energy of the conformationally heterogeneous denatured state ensemble. The other possibility is that the free energy of the transition state increases upon mutation. This could come about if the transition state contains unfavorable terms, such as a partially-buried serine side chain that is limited in its ability to form hydrogen bonds, or if native-like stabilizing interactions are simply missing from the transition state. While the nonconservative nature of the substitution introduces some ambiguity into this interpretation, the idea that the aromatic cluster residues are partly involved in the folding nucleus is corroborated by the results of alanine substitutions of the analogous phenylalanines in *Bacillus subtilis* CspB (Jacob et al., 1999). We are unfortunately unable to draw a direct comparison to the CspB study since the folding kinetics of the alanine mutants were studied at a much higher temperature.

Phe $\rightarrow$ Leu is a relatively conservative substitution that removes an aromatic ring but preserves some bulk. Phenylalanine and leucine have essentially the same number of accessible side-chain torsion angles. It could even be argued that this substitution is more appropriate for measuring transition state effects than a Phe $\rightarrow$ Ala substitution, because the latter results in a large change in side-chain volume that could cause structural rearrangements resulting in altered interactions being made by residues in the vicinity of the change. Also, despite the difference in size between these two side chains, phenylalanine and leucine actually have comparable free energies of transfer into octanol (Fauchere & Pliska, 1983) and should have similar energetics in the unfolded state.

The substitution of surface phenylalanines 20 and 31 to leucines leads to somewhat unusual changes in the folding kinetics of CspA (FLF and FFL in Tables 1 and 2). These changes cause 1.5–2-fold refolding rate *increases* that are larger than the increases in the unfolding rates and lead to large, negative ϕ_f -values. Combining these two mutations (FLL) leads to a further cumulative increase in the refolding rate of CspA. These rate accelerations lead us to speculate that the organization of leucine residues is more rapid than the organization of the more bulky phenylalanines residues. In other words, the leucine mutations may lower the entropic barrier to folding, perhaps because the leucines can make equivalent stabilizing interactions in a variety of side-chain orientations in the folding transition state ensemble. An alternative explanation for the negative ϕ_f -values we observe in the FFL and FLF mutants (Table 2) is that the phenylalanines of the wild-type protein make nonnative destabilizing interactions in the folding transition state, accounting for the higher free energy of this state in comparison to that in the leucine mutants. We believe that this is less likely in

light of the effects of serine substitutions in CspA and alanine substitutions in CspB. Detailed kinetic studies of the folding of the leucine mutants across a range of temperatures may shed more light on their energetics of folding.

By characterizing the folding kinetics of the complete set of the Phe $\rightarrow$ Leu mutants, we are able to determine the source of the energetic coupling between the phenylalanines that we had observed previously (Hillier et al., 1998). From the mutant cycle analysis shown in Figure 3, it appears that all of the observed cooperativity results from changes in the free energy differences between the folded and the transition state ensembles of the mutants. Thus, when considering the folding of CspA, the cooperative interactions between the phenylalanines are not present until the after transition state has been reached: the aromatic groups act in concert to stabilize the folded state of CspA against unfolding, but do not assist in folding. Although we observe cumulative refolding rate accelerations observed upon mutating the phenylalanines to leucines, the increases are merely additive, and each residue's contribution is independent of another's.

Even though the mutations described here appear to affect the free energy of the folding transition state, they do not seem to change the native-like character of the transition state ensemble as reflected in the value of $\alpha^\ddagger$ (Table 2). Only the fast-folding double mutant FLL has a modest but significant change in the solvent accessibility of its transition state toward the unfolded state. This shift reinforces the involvement of residues F20 and F31 in the transition state structure of CspA. Apart from this slight change, however, the native-like transition state thus far seems to be relatively impervious to mutation and is conserved among all members of the cold shock family whose folding has been studied thus far (Perl et al., 1998; Reid et al., 1998; Schindler et al., 1999).

Cold shock protein evolution

Proteins are clearly constrained by evolution to function and to be stable. It has also been speculated that evolution selects for proteins that fold rapidly (Levinthal, 1968; Sali et al., 1994; Mirny et al., 1998), though this hypothesis is under debate (Kim et al., 1998). Overall, our results implicate the same group of residues in CspA as contributing to nucleic acid binding activity, protein stability, and, in part, to directing protein folding. Therefore, several separate evolutionary constraints must act simultaneously on the same group of residues in CspA to preserve function, stability, and folding. CspA is not unique in this respect: some residues in the α -helical protein, CheY, are involved in nucleating folding as well as in binding (López-Hernández & Serrano, 1996). Nonetheless, mutation of the aromatic cluster residues in CspA to leucines reduces stability and nucleic acid binding function (Hillier et al., 1998), but accelerates folding. So, when push came to shove, stability and function were favored over faster folding in the evolution of the cold shock proteins.

Mirny and Shakhnovich (1999) have predicted which residues form the folding nucleus in CspA by considering which positions are consistently conserved in families of nonhomologous proteins sharing the same OB fold. The solvent-exposed aromatic cluster residues were not predicted to be part of the nucleus because, even though they are highly conserved in the cold shock family as well as in very distantly-related nucleic acid binding OB fold proteins, they are not universally conserved in this fold. While we have yet to investigate whether other residues in CspA contribute to folding nucleation, the results presented here show that in CspA, the aro-

matic cluster residues, particularly F20 and F31, do participate in the folding transition state ensemble.

The cold shock proteins are not particularly stable, with the mesophilic members of this family having unfolding free energies on the order of 3 kcal/mol. It has been speculated that this marginal stability has a regulatory role (Schindler et al., 1995; Reid et al., 1998; Mirny & Shakhnovich, 1999). One test of this hypothesis would be to examine the folding of cold shock domains of entirely different function (e.g., transcription regulation from a eukaryotic organism). Such experiments could help to determine whether the marginal stability and fast folding properties of the bacterial cold shock proteins are properties of the fold or consequences of regulation.

Implications for β -sheet formation

The primary question that has driven our investigation of the folding properties of CspA is: What are the sequence determinants that code for β -sheet formation and stability? Some of the factors thought to be important are: (1) strong intrinsic β -sheet propensity for residues within strands; (2) specific side-chain–side-chain interactions between residues on neighboring strands; (3) strong turn signals; and (4) an alternating polar/nonpolar sequence pattern within strands. We will now consider how our findings on the aromatic cluster mutants of CspA relate to each of these factors.

β -sheet propensity

Of the three residue types examined, phenylalanine and serine have high β -forming propensities and leucine has a medium to low β -forming propensity (Minor & Kim, 1994a, 1994b; Smith et al., 1994). Our results on the collection of CspA mutants suggest that β -sheet propensity is not the dominant driving force for fast folding, since the mutants containing leucine, the worst β -former of the three residues fold the fastest. Intrinsic β -sheet propensity may be more important for maintaining protein stability, though the local context clearly influences the magnitude of that effect, since leucine mutants are still more stable than serine mutants (Hillier et al., 1998) in spite of the lower β -sheet propensity of leucine.

Specific side-chain interactions

Several statistical studies of side-chain–side-chain interactions in antiparallel β -sheets have been performed (Wouters & Curmi, 1995; Hutchinson et al., 1998; Zhu & Braun, 1999). Even though the crystallographic database has grown in recent years, interactions involving rarer amino acids can still be too infrequent to be statistically significant. For all side chains, there is still insufficient data to carefully compute pairing statistics in a way that normalizes for the environment in which a pair is found as well as for whether the backbones of the pair are hydrogen bonded or non-hydrogen bonded. With these limitations in mind, we computed the cumulative pairing propensities of the leucine and serine mutants in this study using the pairing statistics computed by Wouters and Curmi (1995) by comparing the interactions made in the different mutants. We found no correlation between the complementarity of residue pairs and either folding rate or stability in our collection of mutants. By far, the best collection of residue pairs is found in mutant SFF even though this is the least stable and second slowest folding mutant of the set. Wild-type CspA, which is most stable and folds rapidly, has the third worst side-chain interaction propensity. These results imply that either the empirical side-chain pairing propensities are still very inaccurate, or that residue pairing

in β -sheets is of a highly context-dependent nature (Zaremba & Gregoret, 1999).

Strong turn signals

Other studies have suggested that certain turns may nucleate the folding of β -sheet proteins (Grantcharova et al., 1998; Martinez et al., 1998b; Martinez & Serrano, 1999; Riddle et al., 1999). The mutants described here do not really probe the importance of turns to β -sheet formation and stability. We only note that mutations to F18, which forms the first hydrogen-bonding interaction bridging a β -turn, have the smallest effect on the free energy of the transition state ensemble.

Polar/nonpolar pattern

Finally, with respect to the alternating polar/nonpolar pattern characteristic of β -strands, it is interesting that in wild-type CspA and the mutants studied here, fast folding is correlated with having nonpolar residues at positions 18, 20, and 31. These residues are all on the surface of the protein. Therefore, in the fast folding variants, β -strands 2 and 3 (on which these residues lie) do not have the alternating polar/nonpolar pattern typical of β -sheets. Furthermore, at least two of these surface residues (20 and 31) appear to take part in nucleating folding. This suggests that in CspA, a polar/nonpolar pattern to β -strands is not the dominant factor that dictates fast folding. Introduction of such a pattern into this region with serine residues also destabilizes rather than stabilizes CspA (Hillier et al., 1998).

In consideration of these factors, we propose that hydrophobic collapse drives fast folding in the cold shock proteins and that both surface and buried nonpolar residues contribute to the collapse. While substituting buried, hydrophobic residues with polar ones is clearly energetically unfavorable (Lim et al., 1992; Blaber et al., 1993), substitution of surface residues usually has little effect on protein stability. It may be therefore that the polar/nonpolar patterns observed in both helical and sheet proteins are a consequence of the requirement for globular proteins to be soluble, rather than a determinant of folding kinetics or thermodynamics. In fact, globular proteins with a random sequence of glutamines, leucines, and arginines are *ultrastable* when the fraction of nonpolar (leucines) residues is higher than normally observed in globular proteins (Davidson & Sauer, 1994; Davidson et al., 1995), suggesting that having too many nonpolar residues is stabilizing rather than destabilizing. Curiously, a recent statistical analysis has shown that strictly alternating polar/nonpolar patterns are actually very rare in protein sequences and may actually promote untoward protein aggregation (Broome & Hecht, 2000).

In summary, our results support the hypothesis that nonpolar clusters on globular protein surfaces stabilize proteins and can lead to faster folding (Tisi & Evans, 1995). The implications of these findings are that it may be possible to accelerate the folding (thereby disfavoring off-pathway aggregation reactions) or increase the stability of biomedically-important recombinant proteins and reagents through the introduction of small nonpolar clusters onto their surfaces.

Materials and methods

Analytical grade urea was purchased from ICN Biomedicals (Costa Mesa, California). Buffer components and other purification reagents were of analytical grade and purchased from Fisher Scientific, Inc. (Springfield, New Jersey). Ni-NTA resin used to purify

CspA and mutants was purchased from Qiagen (Hilden, Germany). Proteins were expressed and purified as described previously (Hillier et al., 1998). Protein concentration was determined using a UV-Visible spectrophotometer. The extinction coefficients were calculated to be 8,437 M⁻¹cm⁻¹ for wild-type CspA, 8,237 M⁻¹cm⁻¹ for the single mutants, 8,037 M⁻¹cm⁻¹ for the double mutants, and 7,837 M⁻¹cm⁻¹ for the triple mutant (Pace et al., 1995).

Stopped-flow fluorescence

All folding and unfolding reactions were performed at 25 °C using a Biologic SFM2 stopped flow apparatus in fluorescence mode. The Biokine 3.0 software (Molecular Kinetics, Pullman, Washington) was used for data acquisition. Unfolding was initiated by rapid dilution of native protein in 50 mM potassium phosphate, 100 mM KCl, pH 7.0 (Buffer A) with the appropriate ratio of denaturing buffer containing different concentrations of urea in the same buffer. The refolding reactions were performed by dilution of unfolded protein in buffer A with 7 M urea into Buffer A with or without urea to give the desired final condition. Excitation was at 280 nm and emission was measured through a 325 nm cutoff filter. Final protein concentrations were between 1–5 μM. At least 10 transients were collected and averaged for each concentration, and these data were fit to a mono-exponential time course using non-linear least-squares regression as implemented by Kaleidagraph (Synergy Software, Reading, Pennsylvania).

The rate constants for refolding (k_f) and unfolding (k_u) and the dependences of these rates on urea concentration (m_f and m_u) were obtained from fits to Equation 1. The fits were repeated leaving out each data point in turn. The average values resulting from these fits along with errors representing the range of values obtained in the fits are reported in Table 1.

$$\lambda = (k_f + k_u) \\ = k_f^{\text{H}_2\text{O}} \cdot \exp(m_f \cdot [\text{urea}]) + k_u^{\text{H}_2\text{O}} \cdot \exp(m_u \cdot [\text{urea}]). \quad (1)$$

The solvent accessibility of the transition state $\alpha^\ddagger$ was computed from the denaturant dependences of the unfolding and refolding rates on urea concentration according to Equation 2 (Tanford, 1970; Chen et al., 1989):

$$\alpha^\ddagger = \frac{m_f}{m_f - m_u} = -\frac{m_f}{m_{eq}} \cdot RT. \quad (2)$$

Φ-Value analysis

The contribution of a residue to the folding transition state free energy was computed using ϕ -value analysis (Fersht, 1995; Matouschek et al., 1995). ϕ_f is the ratio of the change in free energy of the transition states with respect to the unfolded state, and the change in free energy between the folded and unfolded states (Equation 3). k_f is the refolding rate in the absence of denaturant and ΔG_U is the free energy of unfolding. In our analysis, ϕ_f is computed in the refolding direction:

$$\phi_f = \frac{\Delta \Delta G^\ddagger}{\Delta \Delta G_U} = \frac{R \cdot T \cdot \ln \left(\frac{k_f^{\text{WT}}}{k_f^{\text{Mut}}} \right)}{\Delta G_U^{\text{WT}} - \Delta G_U^{\text{Mut}}}. \quad (3)$$

ΔG_u s computed from equilibrium analysis (Hillier et al., 1998) as well as from kinetic analysis were used to provide alternative, corroborating calculations of ϕ_f .

Thermal denaturation

Thermal denaturation for each protein was followed using CD spectroscopy by measuring the change in ellipticity at 222 nm with increasing temperature. The protein samples were in Buffer A at a concentration of ~50 μM. The temperature was controlled using a programmable water bath (Neslab) set to heat at a rate of 0.33 °C/min. The experimental data, ellipticity vs. temperature, were fit to Equation 4 using Kaleidagraph to obtain the midpoint of thermal denaturation (T_m). ΔG is given by Equation 5 and $y_f(T)$ and $y_u(T)$ represent the folded and unfolded baselines as a function of temperature, respectively.

$$\theta = \frac{\exp \left(-\frac{\Delta G}{RT} \right)}{1 + \exp \left(-\frac{\Delta G}{RT} \right)} [y_f(T) - y_u(T)] + y_u(T) \quad (4)$$

$$\Delta G = \frac{T_m - T}{T_m} \Delta H - (T_m - T) \Delta C_p + T_m \Delta C_p \ln \left(\frac{T_m}{T} \right). \quad (5)$$

Mutant cycle analysis

Double mutant cycles were constructed for each of the residue interactions as discussed previously (Hillier et al., 1998) with the following exceptions. The Gibbs free energy of unfolding was computed for each protein using Equation 6 and the respective extrapolated rate constants. In the case of the transition state free energies, Equations 7 and 8 were used. The differences in free energies ($\Delta \Delta G$) between each protein (in the direction of the arrows, for example, in Equation 9 we compute the $\Delta \Delta G$ for the FFF → FFL mutation) were calculated from the free energies of each mutant. The interaction energies between the three possible pairs of residues (ΔG_{int} , equivalent to $\Delta \Delta G$, Equation 10) were determined from the free energy difference between any of the parallel arrows (Horovitz & Fersht, 1992).

$$\Delta G_U = -RT \ln \left(\frac{k_u^{\text{H}_2\text{O}}}{k_f^{\text{H}_2\text{O}}} \right) \quad (6)$$

$$\Delta G_u^\ddagger = -RT \ln k_u^{\text{H}_2\text{O}}. \quad (7)$$

$$\Delta G_f^\ddagger = -RT \ln k_f^{\text{H}_2\text{O}}. \quad (8)$$

$$\Delta \Delta G_{\text{FFF} \rightarrow \text{FLF}}^\ddagger = \Delta G_{\text{FLF}}^\ddagger - \Delta G_{\text{FFF}}^\ddagger \quad (9)^1$$

$$\Delta G_{int} = \Delta(\Delta \Delta G^\ddagger) = \Delta \Delta G_{\text{FFL} \rightarrow \text{FLL}}^\ddagger - \Delta \Delta G_{\text{FFF} \rightarrow \text{FLF}}^\ddagger. \quad (10)^1$$

¹The interaction energy between F20 and F31 is shown here as an example. The other interaction energies were calculated in an analogous manner.

Acknowledgments

We wish to thank Kevin Plaxco for helpful comments on the manuscript and Brian Hillier for assistance with protein purification. This work was supported by NIH grant GM52885, including a minority student award supplement, an award from the University of California Biotechnology Research and Education Program, and an American Chemical Society Petroleum Research Fund starter grant.

References

- Abkevich VI, Gutin AM, Shakhnovich EI. 1994. Specific nucleus as the transition state for protein folding: Evidence from the lattice model. *Biochemistry* 33:10026–10036.
- Blaber M, Lindstrom JD, Gassner N, Xu J, Heinz DW, Matthews BW. 1993. Energetic cost and structural consequences of burying a hydroxyl group within the core of a protein determined from Ala → Ser and Val → Thr substitutions in T4 lysozyme. *Biochemistry* 32:11363–11373.
- Broome BM, Hecht MH. 2000. Nature disfavors sequences of alternating polar and non-polar amino acids: Implications for amyloidogenesis. *J Mol Biol* 296:961–968.
- Carson M. 1991. RIBBONS 2.0. *J Appl Crystallogr* 24:958.
- Chen B, Baase WA, Schellman JA. 1989. Low-temperature unfolding of a mutant of phage T4 lysozyme. 2. Kinetic investigations. *Biochemistry* 28:691–699.
- Davidson AR, Lumb KJ, Sauer RT. 1995. Cooperatively folded proteins in random sequence libraries. *Nat Struct Biol* 2:856–864.
- Davidson AR, Sauer RT. 1994. Folded proteins occur frequently in libraries of random amino acid sequences. *Proc Natl Acad Sci USA* 91:2146–2150.
- Fauchere J-L, Pliska V. 1983. Hydrophobic parameters π of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. *Eur J Med Chem* 18:369–375.
- Fersht AR. 1995. Characterizing transition states in protein folding: An essential step in the puzzle. *Curr Opin Struct Biol* 5:79–84.
- Fersht AR, Matouschek A, Serrano L. 1992. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 224:771–782.
- Grantcharova VP, Riddle DS, Santiago JV, Baker DB. 1998. Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat Struct Biol* 5:714–720.
- Hillier BJ, Rodriguez HM, Gregoret LM. 1998. Coupling of protein stability and protein function in *E. coli* CspA. *Fold Des* 3:87–93.
- Horovitz A, Fersht AR. 1992. Co-operative interactions during protein folding. *J Mol Biol* 224:733–740.
- Hutchinson GE, Sessions RB, Thornton JM, Woolfson DN. 1998. Determinants of strand register in antiparallel β -sheets of proteins. *Protein Sci* 7:2287–2300.
- Itzhaki LS, Otzen DE, Fersht AR. 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* 254:260–288.
- Jackson SE. 1998. How do small single-domain proteins fold? *Fold Des* 3:R81–R91.
- Jacob M, Holtermann G, Perl D, Reinstein J, Schindler T, Geeves MA, Schmid FX. 1999. Microsecond folding of the cold shock protein measured by a pressure-jump technique. *Biochemistry* 38:2882–2891.
- Kim DE, Gu H, Baker D. 1998. The sequences of small proteins are not extensively optimized for rapid folding by natural selection. *Proc Natl Acad Sci USA* 95:4982–4986.
- Levinthal C. 1968. Are there pathways for protein folding? *J Chim Phys* 65:44–45.
- Lim WA, Farruggio DC, Sauer RT. 1992. Structural and energetic consequences of disruptive mutations in a protein core. *Biochemistry* 31:4324–4333.
- López-Hernández E, Serrano L. 1996. Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2. *Fold Des* 1:43–55.
- Martinez JC, Pisabarro MT, Serrano L. 1998a. Obligatory steps in protein folding and the conformational diversity of the transition state. *Nat Struct Biol* 5:721–729.
- Martinez JC, Pisabarro MT, Serrano L. 1998b. Obligatory steps in protein folding and the conformational diversity of the transition state. *Nat Struct Biol* 5:721–729.
- Martinez JC, Serrano L. 1999. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat Struct Biol* 6:1010–1016.
- Matouschek A, Otzen DE, Itzhaki LS, Jackson SE, Fersht AR. 1995. Movement of the position of the transition state in protein folding. *Biochemistry* 34:13656–13662.
- Minor DL, Kim PS. 1994a. Context is a major determinant of β -sheet propensity. *Nature* 371:264–267.
- Minor DL, Kim PS. 1994b. Measurement of the beta-sheet-forming propensities of amino acids. *Nature* 367:660–663.
- Mirny LA, Abkevich VI, Shakhnovich EI. 1998. How evolution makes proteins fold quickly. *Proc Natl Acad Sci USA* 95:4976–4981.
- Mirny LA, Shakhnovich EI. 1999. Universally conserved positions in protein folds: Reading evolutionary signals about stability, folding kinetics and function. *J Mol Biol* 291:177–196.
- Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. 1995. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* 4:2411–2423.
- Perl D, Welker C, Schindler T, Schroeder K, Maraheil MA, Jaenicke R, Schmid FX. 1998. Conservation of rapid two-state folding in mesophilic, thermophilic, and hyperthermophilic cold shock proteins. *Nat Struct Biol* 5:229–235.
- Reid KL, Rodriguez HM, Hillier BJ, Gregoret LM. 1998. Stability and folding properties of a model β -sheet protein, *Escherichia coli* CspA. *Protein Sci* 7:470–479.
- Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I, Baker D. 1999. Experiment and theory highlight role of native state topology in SH3 folding. *Nat Struct Biol* 6:1016–1024.
- Sali A, Shakhnovich E, Karplus M. 1994. How does a protein fold? *Nature* 369:248–251.
- Schellman JA. 1978. Solvent denaturation. *Biopolymers* 17:1305–1322.
- Schindelin H, Jiang W, Inouye M, Heinemann U. 1994. Crystal structure of CspA, the major cold shock protein of *Escherichia coli*. *Proc Natl Acad Sci USA* 91:5119–5123.
- Schindler T, Graumann PL, Perl D, Ma S, Schmid FX, Marahiel MA. 1999. The family of cold shock proteins of *Bacillus subtilis*. Stability and dynamics in vitro and in vivo. *J Biol Chem* 274:3407–3413.
- Schindler T, Herrler M, Maraheil MA, Schmid FX. 1995. Extremely rapid protein folding in the absence of intermediates. *Nat Struct Biol* 2:663–673.
- Smith CK, Regan L. 1995. Guidelines for protein design: The energetics of β -sheet side chain interactions. *Science* 270:980–982.
- Smith CK, Withka JM, Regan L. 1994. A thermodynamic scale for β -sheet forming tendencies of the amino acids. *Biochemistry* 33:5510–5517.
- Tanford C. 1970. Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* 24:1–95.
- Tisi LC, Evans PA. 1995. Conserved structural features on protein surfaces: Small exterior hydrophobic clusters. *J Mol Biol* 249:251–258.
- Wouters MA, Curmi PMG. 1995. An analysis of side chain interactions and pair correlations within antiparallel β -sheets: The differences between backbone hydrogen-bonded and non-hydrogen-bonded residue pairs. *Proteins* 22:119–131.
- Zaremba SM, Gregoret LM. 1999. Context-dependence of amino acid residue pairing in antiparallel β -sheets. *J Mol Biol* 291:463–479.
- Zhu H, Braun W. 1999. Sequence specificity, statistical potentials, and three-dimensional structure prediction with self-correcting distance geometry calculations of beta-sheet formation in proteins. *Protein Sci* 8:326–342.