Critical Side-Chain Interactions at a Subunit Interface in the Arc Repressor Dimer†

Marcos E. Milla‡ and Robert T. Sauer*

Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Received September 19, 1994; Revised Manuscript Received November 29, 1994⊗

ABSTRACT: In the Arc repressor dimer, the side chains of Ile37 and Val41 in α-helix B pack against each other and against the symmetry-related side chains of Ile37′ and Val41′ in α-helix B′ to form part of the hydrophobic core and the dimer interface. Following combinatorial mutagenesis of these positions, only the wild-type combination of hydrophobic residues was recovered as a fully active protein, and only a few conservative replacements were recovered as stably folded or partially active proteins. Equilibrium and kinetic studies of the folding of purified mutants show that the δ-CH_3 groups of Ile37 and Ile37′ contribute approximately 2 kcal/mol of dimer to protein stability and are involved in interactions that are only partially formed in the transition state for protein folding. Alanine substitution at either position 37 or 41 results in proteins which differ from wild type in being monomeric at a concentration of 10 μM, having reduced secondary structure, having solvent-exposed tryptophans, and showing non-cooperative thermal and urea denaturation transitions. These mutants appear to exist in a physiologically denatured state that is similar in many ways to the molten globule state.

Understanding how an amino acid sequence determines the three-dimensional structure and stability of a protein is an important but extremely challenging problem (Matthews, 1993). One complication is that structural information is distributed unevenly throughout the sequence. Some residues are critical in determining structure or stability, while others play no significant role in these processes. We are studying how the sequence of Arc repressor determines the structure and stability of the protein. Arc is a 53-residue, homodimeric, DNA-binding protein encoded by phage P22 (Susskind & Youderian, 1983; Vershon et al., 1985, 1987; Knight et al., 1989). The NMR and crystal structures of the Arc dimer show intertwined monomers, each consisting of a β-strand and two α-helices (Breg et al., 1990; Bonvin et al., 1994; C. Kissenger, U. Obeysekare, L.J. Keefe, B. E. Raumann, R. T. Sauer, and C. O. Pabo, in preparation). The β-strands form an antiparallel β-sheet in the dimer, which packs together with the four helices to form a roughly globular structure. The cocrystal structure of Arc, bound as a tetramer to operator DNA, is also known (Raumann et al., 1994).

Random and directed mutagenesis have been used to probe the importance of individual residues of Arc in determining stable folding (Vershon et al., 1986; Bowie & Sauer, 1989a; Milla et al., 1994). For example, in a scanning mutagenesis study, five buried residues—Val22, Glu36, Ile37, Val41, and Phe45—were identified as playing especially important roles because alanine substitution mutations at these positions were the most severely destabilizing (Milla et al., 1994). Moreover, because the ten side chains affected by these five mutations cluster together in the Arc dimer, it was proposed that these mutations may identify a primitive core that serves as a foundation for Arc folding.

In this paper, we focus on two of the critical Arc residues, Ile37 and Val41. As shown in Figure 1, the Ile37 and Val41 side chains are on the same surface of helix B and pack against each other and against the Ile37′ and Val41′ side chains from the symmetry-related helix B′. These packing interactions are an integral part of the hydrophobic core and constitute the central portion of the dimer interface. Mutagenic studies coupled with biophysical characterization of a subset of the mutant proteins allow several conclusions: (i) Positions 37 and 41 are intolerant toward most substitutions; only the wild-type combination of hydrophobic residues yields full biological function, and most combinations of other residues at these positions appear to be quite destabilizing. (ii) The δ-CH_3 groups of Ile37 and Ile37′ contribute 1.5−2 kcal/mol of dimer to protein stability and are involved in interactions at or before the transition state for protein folding and also subsequent to this step. (iii) Proteins with alanines at either or both positions have properties expected for molten globules and may provide models for physiologically denatured Arc.

MATERIALS AND METHODS

Mutagenesis. The codons for Arc residues 37 and 41 were mutagenized by synthesis of an *Eco*RI/*Hin*dIII cassette (encoding Arc residues 34−44; Bowie & Sauer, 1989a) containing three bases (T, A, and G) at the first codon position, two bases (T and C) at the second position, and two bases (G and C) at the third position on the sense strand of the cassette. On the antisense strand of the cassette, complementary bases were used at the randomized positions (e.g., A and G at the second codon position). This mutagenic strategy results in codons encoding Ala, Phe, Ile, Leu, Met, Ser, Thr, and Val. Cassettes were ligated to the large *Eco*RI/*Hin*dIII backbone fragment of pSA500, a plasmid which encodes Arc with the st5 C-terminal extension KNQHE (Milla et al., 1993), and the mixture was transformed into

† This work was supported by NIH Grant AI-15706 and by a postdoctoral fellowship to M.E.M. from the Jane Coffin Childs Fund for Medical Research.

* Corresponding author.

‡ Current address: Department of Biochemistry, Duke University Medical Center, Durham, NC 27710.

⊗ Abstract published in *Advance ACS Abstracts,* February 15, 1995.

0006-2960/95/0434-3344$09.00/0

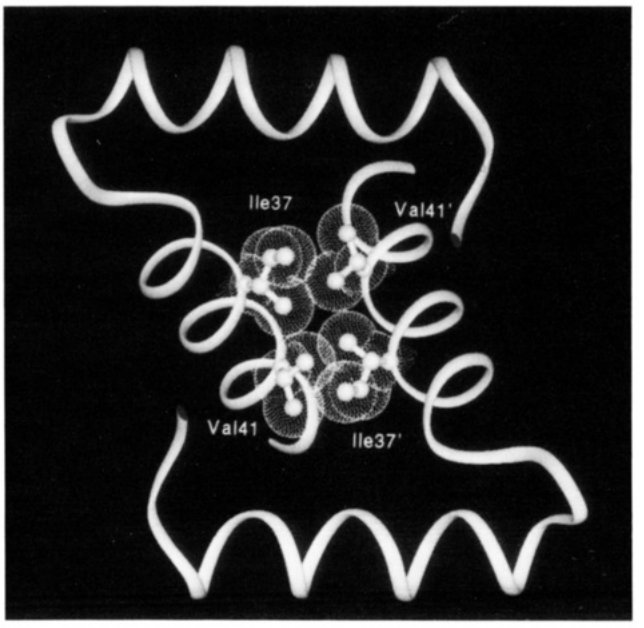

FIGURE 1: Packing interactions between the side chains of Ile37, Ile37′, Val41, and Val41′ in the Arc repressor dimer. The side chains are shown in ball-and-stick representation with van der Waals surfaces. The polypeptide backbones of residues 12−46 of each Arc monomer are shown as a ribbon trace. The view is from the β-sheet of Arc, most of which has been removed for clarity.

Escherichia coli strain UA2F cells by electroporation. In addition to the *arc* gene, pSA500 also contains a fusion of the Arc-repressible P_{ant} promoter to a dominant gene encoding streptomycin sensitivity (Bowie & Sauer, 1989b). Strain UA2F contains an integrated $\lambda imm21$ prophage bearing a transcriptional fusion of the P_{ant} promoter to the *cat* gene (Vershon et al., 1986) and is *strA thi*$^-$ *his*$^-$ *lac*Z$^-$ *lac*Y$^+$ *sup*^{0} *rec*A$^-$ F′(*lac*I^Q *lac*Z::Tn5 [*kan*^{R}] *pro*$^+$). The P_{ant}-*str*^{S} and P_{ant}-*cat* fusions allow antibiotic selections for and against Arc repressor function *in vivo*. In the first selection, Arc activity is required for the cells to survive in the presence of streptomycin (Bowie & Sauer, 1989b; Mossing et al., 1991). For these assays freshly transformed cells were plated onto LB agar plates containing 100 μg/mL ampicillin, 25 μg/mL kanamycin, 50 μg/mL streptomycin, and 2 μg/mL IPTG. In the second selection, cells bearing active Arc variants do not survive in the presence of chloramphenicol (Vershon et al., 1986). LB agar plates containing 100 μg/mL ampicillin, 25 μg/mL kanamycin, and 75 μg/mL chloramphenicol, with or without 2 μg/mL IPTG, were used for these assays. The *arc* genes from colonies surviving either antibiotic selection were sequenced to determine the identity of mutations at positions 37 and 41.

Protein Expression and Purification. The steady-state expression levels of mutant proteins in strain UA2F/pSA500 were assayed by SDS-polyacrylamide gel electrophoresis of crude cell lysates as described (Milla et al., 1993). In each case tested, expression levels were found to be similar to that of wild-type Arc-st5 or were undetectable.

To facilitate expression and affinity purification, selected mutations were subcloned from the *arc-st5* gene of pSA500 into the *arc-st11* genes of pSA700 (Milla et al., 1993) or pET800 (Milla et al., 1994). pET800 is a vector in which Arc's expression is controlled by the T7 promoter and terminator sequences (Rosenberg et al., 1987; Studier & Moffatt, 1986). Arc-st11 contains the C-terminal extension H_6KNQHE (Milla et al., 1993). Wild-type Arc-st11 and the variants were purified from strains UA2F/pSA700 or X90 (λDE3)/pET800 by chromatography on Ni^{2+}-NTA agarose (Qiagen, Inc.) and SP-Sephadex (Pharmacia) as previously described (Brown et al., 1994). Protein concentrations in monomer equivalents per liter were determined using an extinction coefficient at 280 nm of 6756 M^{-1} cm^{-1} (Brown et al., 1990).

Biophysical Methods. Unless noted, all experiments were performed at 25 °C in a standard buffer containing 50 mM Tris-HCl (pH 7.5), 250 mM KCl, and 0.2 mM EDTA. Equilibrium, kinetic, and sedimentation data were fitted by nonlinear least-squares procedures using a Macintosh version of the program NonLin (Johnson & Frasier, 1985; Brenstein, 1989). The statistics subroutine of the program was used to determine 67% confidence limits for fitted parameter values, and these limits were propagated in subsequent calculations as described (Bevington, 1969).

For urea denaturation studies, Arc at a concentration of 10 μM in buffer was placed in a thermostated cuvette at 25 °C, and the urea concentration was increased in an iterative fashion by addition of aliquots of a 10 μM solution of Arc in buffer plus 10 M urea. After each addition of the urea solution, the sample was mixed and equilibrated for at least 40 s, and the circular dichroism (CD) ellipticity at 234 nm was averaged for 40 s. For thermal melts, Arc (10 μM) was heated in 1 °C steps, the sample was allowed to equilibrate for 1 min at each temperature, and the ellipticity at 222 nm was averaged for 30 s. Both denaturation reactions were greater than 95% reversible and were fit to a two-state model in which two denatured monomers are in equilibrium with the native dimer [for details, see Bowie and Sauer (1989c) and Milla et al. (1993)]. For thermal melts, values of t_m, ΔH_u, and the slopes and intercepts of the native and denatured baselines were obtained from the fitting procedure, assuming a ΔC_p value of 1.31 kcal/mol (Milla et al., 1994). Changing ΔC_p by ±0.1 kcal/mol did not alter the derived values of t_m or ΔH_u significantly. For urea melts, the values of ΔG_u in the absence of urea (at a standard-state protein concentration of 1 M), m, and the slopes and intercepts of the native and denatured baselines were obtained from the fitting procedure. C_m, the urea concentration at which 50% denaturation occurs, was calculated as

$$C_m = \frac{\Delta G_u + RT \ln(P_t)}{m} \quad (1)$$

where P_t is the total protein concentration in monomer equivalents. For mutants, values of the unfolding free energy ΔG_u at temperature T (here, the wild-type t_m) were calculated from thermal denaturation parameters using the equation

$$\Delta G_u = \Delta H^\circ - T\Delta S^\circ + \Delta C_p[T - T^\circ - T \ln(T/T^\circ)] \quad (2)$$

where ΔH° and ΔS° are the enthalpy and entropy of unfolding/dissociation at reference temperature T° (here, the mutant t_m) (Becktel & Schellman, 1987).

The binding of 1-anilino-8-naphtalenesulfonate (ANS) to Arc and mutants was assayed in standard buffer at 25 °C by monitoring changes in the fluorescence emission of the dye at 480 nm after excitation at 400 nm. For these experiments, the protein concentration was 50 μM and the ANS concentration was 5−300 μM.

The rates of refolding and unfolding of Arc and mutants were measured by using a stop-flow instrument (Applied Photophysics DX17-MV) to monitor changes in tryptophan fluorescence (Milla & Sauer, 1994). Refolding was initiated by mixing acid-denatured protein (in 10 mM phosphoric acid (pH 2.0), 250 mM KCl, and 0.2 mM EDTA) with an equal volume of refolding buffer (100 mM Tris-HCl (pH 8.0), 250 mM KCl, and 0.2 mM EDTA). Following mixing, the pH was 7.5 and the protein concentration was 4−5 μM. Unfolding reactions were initiated by mixing Arc in standard buffer with buffer containing urea. In these experiments, the final protein concentration was 5 μM and the final urea concentration was between 3 and 6 M. Kinetic data for refolding and unfolding were fit as described by Milla and Sauer (1994).

The oligomeric state of Arc mutants was determined in 50 mM Tris-HCl (pH 7.5), 250 mM KCl, and 0.2 mM EDTA by sedimentation equilibrium centrifugation at speeds of 27 000 and 39 000 rpm using a Beckman-XLA analytical ultracentrifuge. Sedimentation data were used to determine average molecular weights (M_a) by fitting to the function

$$\mathrm{Abs}(x) = \mathrm{Abs}(x_f)\exp[M_a(\omega^2/2RT)(1 - \nu\varrho)(x^2 - x_f^2)] \quad (3)$$

where x and x_f are the radius and an arbitrary reference radius, respectively, Abs(x) and Abs(x_f) are the absorbances at x and x_f, ν is the partial specific volume, ϱ is the density of the solution, ω is the angular velocity, R is the gas constant, and T is the temperature (Laue et al., 1992). The partial specific volume was calculated from the protein sequence as described (Laue et al., 1992). For the alanine mutants, sedimentation data were also used to estimate equilibrium constants for dimer unfolding/dissociation using the function

$$\mathrm{Abs}_{(x)} = A\exp[M_i(\omega^2/2RT)(1 - \nu\varrho)(x^2 - x_f^2)] + B\exp[2M_i(\omega^2/2RT)(1 = \nu\varrho)(x^2 - x_f^2)] \quad (4)$$

Here, the definitions are the same as in eq 1, except that M_i is the monomer molecular weight calculated from the protein sequence, A and B are variables that depend on the concentration of monomer (A) and dimer (B) at radius x_f, and the partial specific volumes of the monomer and dimer species are represented by a single value of ν. The equilibrium constant for unfolding/dissociation of the dimer can be calculated as

$$K_u = (2P_t/\mathrm{Abs}_o)/(A^2/B) \quad (5)$$

where P_t is the total monomer concentration and Abs_o is the initial absorbance of the sample before sedimentation.

RESULTS

Tolerance of Positions 37 and 41 to Hydrophobic Substitutions. The side chains of Ile37 and Val41 are completely buried in the hydrophobic core of the wild-type Arc dimer. Although substitutions of polar residues for hydrophobic core side chains are generally destabilizing, it is frequently found that substitutions with other hydrophobic side chains can be accommodated [for a review, see Richards & Lim (1994)]. To test this possibility, the codons for Arc residues 37 and 41 were mutagenized to generate all 64 combinations of six hydrophobic amino acids (Ala, Ile, Leu, Met, Phe, and Val) and two additional residues (Ser and Thr). Although Ser and Thr are somewhat polar, they are found at some buried positions in α-helices where their side-chain hydroxyls can hydrogen bond to the polypeptide main chain (Gray & Matthews, 1984; Blaber et al., 1993a).

Table 1: Recovery of Active and Inactive Sequences after Randomization of Codons 37 and 41[a]

ACTIVE

		41 A	41 S	41 T	41 V	41 I	41 L	41 M	41 F
37	A	–	–	–	–	–	–	–	–
37	S	–	–	–	–	–	–	–	–
37	T	–	–	–	–	–	–	–	–
37	V	–	–	–	**1**	–	–	–	–
37	I	–	–	–	**12**	**2**	**1**	–	–
37	L	–	–	–	**8**	–	–	–	–
37	M	–	–	–	–	–	–	–	–
37	F	–	–	–	–	–	–	–	–

INACTIVE

		41 A	41 S	41 T	41 V	41 I	41 L	41 M	41 F
37	A	3	8	5	4	3	2	1	1
37	S	4	6	3	5	1	1	–	3
37	T	5	3	1	3	1	2	2	4
37	V	4	1	–	**5**	1	2	1	3
37	I	+	–	**2**	–	**1**	**2**	–	–
37	L	–	2	–	**1**	–	–	1	–
37	M	2	1	1	1	–	3	1	1
37	F	–	3	1	1	–	–	–	1

[a] The values listed represent the number of colonies recovered for a particular dipeptide combination following selections for activity and against activity. Dashes represent zero colonies. Colonies indicated by shading had near wild-type levels of protein expression; all others had no detectable expression. The IA37 mutant was not recovered after randomization but was available from a previous study (Milla et al., 1994) and is inactive in both assays

The mutagenized *arc* genes were introduced into *E. coli* UA2F cells, and selections for and against Arc activity were applied in separate experiments. First, *arc* genes encoding active repressors were isolated using a streptomycin resistance selection (Bowie & Sauer, 1989b; Mossing et al., 1991). Of 24 isolates, only wild type (Ile37/Val41) and four conservative single mutants (IL37, IV37, VI41, and VL41) were recovered (Table 1). Each of these proteins was found to be expressed at steady-state levels comparable to wild type. Second, *arc* genes were isolated following a chloramphenicol selection in which cells containing active Arc are killed (Vershon et al., 1986). Here, 48 different combinations were recovered from 113 isolates (Table 1). These include the four single mutants recovered in the streptomycin selection, one additional single mutant and two double mutants (VT41, IV37/VL41, and IV37/VI41) that were found to be expressed at wild type levels, and 42 mutants for which intracellular expression could not be detected. Recovery of the IL37, IV37, VI41, and VL41 variants in both selections suggests

Table 2: Spectral and Oligomeric Properties of Arc-st11 and Mutants

protein	fluorescence[a] (nm) 10 M	fluorescence[a] (nm) 400 μM	Θ_{222} (mdeg) 10 μM	K_u (M)	ΔG_u (kcal/mol dimer)	mol mass[b] (kDa) 10 μM	mol mass[b] (kDa) 400 μM
Arc-st11	335.5		−70.0	$8.5(2.5) \times 10^{-9}$	10.99(0.17)	14.1	
IV37	337.0		−70.6	$2.0(0.5) \times 10^{-7}$	8.99(0.16)	13.4	
VI41	335.5		−71.4	$9.4(3.1) \times 10^{-9}$	10.92(0.20)	14.3	
IV37/VI41	337.0		−71.3	$1.0(0.3) \times 10^{-7}$	9.28(0.17)	14.5	
IA37	351.0	348.0	−27.6	$4.4(0.7) \times 10^{-4}$	4.56(0.09)	8.2	9.9
VA41	349.0	342.5	−25.1	$2.1(0.6) \times 10^{-4}$	4.99(0.18)	7.9	10.9
IA37/VA41	351.5	349.5	−29.7	$7.0(0.8) \times 10^{-4}$	4.28(0.06)	7.4	9.5

[a] Center of mass (±0.25 nm) of the fluorescence emission spectrum from 300 to 400 nm. The standard errors for the ellipticity measurements are roughly ±5%. [b] Average molecular masses calculated from sedimentation equilibrium centrifugation. The monomer molecular weight calculated from the Arc-st11 sequence is 7.67 kDa. K_u and ΔG_u values (±sd) for wild type and the IV37, VI41, and IV37/VI41 mutants are from urea denaturation experiments; K_u and ΔG_u values for the IA37, VA41, and IA37/VA41 mutants were calculated from sedimentation equilibrium centrifugation experiments.

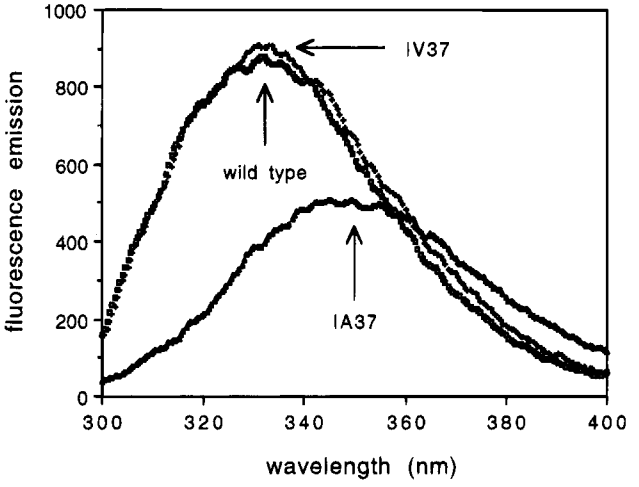

FIGURE 2: Fluorescence emission spectra in arbitrary units of wild-type Arc-st11 and the IV37 and IA37 mutants after excitation at 280 nm (10 μM protein, 25 °C, 50 mM Tris-HCl (pH 7.5), 250 mM KCl, 0.2 mM EDTA).

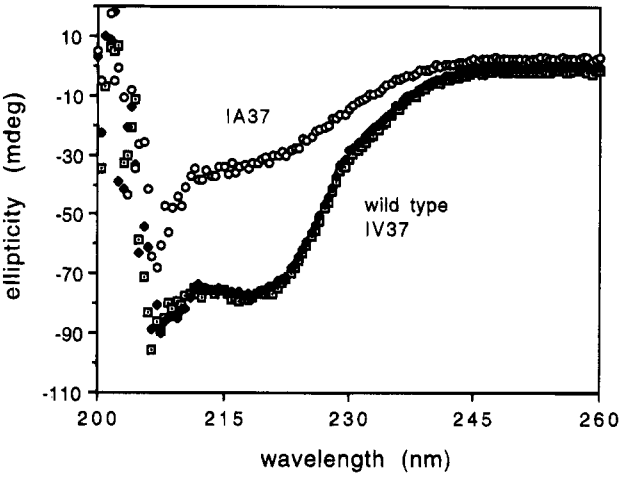

FIGURE 3: Circular dichroism spectra of wild-type Arc-st11 (□) and the IV37 (◆) and IA37 (○) mutants (10 μM protein, 25 °C, 50 mM Tris-HCl (pH 7.5), 250 mM KCl, 0.2 mM EDTA).

that these mutants have partial activities near the detection thresholds of the selections.

Effects of Substitutions on Oligomeric State, Tryptophan Burial, and Secondary Structure. To characterize the effects of the position 37 and 41 mutations on the properties of Arc, three conservative mutants (IV37, VI41, and IV37/VI41) and three more extreme mutants (IA37, VA41, and IA37/VA41) were chosen for analysis. For the latter set of alanine mutants, expression levels adequate for purification could only be obtained by recloning the mutations into the st11 C-terminal tail background (H_6KNQHE; Milla et al., 1993), which provides greater protection from intracellular degradation, and using the T7 promoter to achieve very high levels of transcription. The addition of C-terminal tail sequences, such as st5 or st11, to Arc does not affect the thermodynamic stability of the folded protein (Bowie & Sauer, 1989b; Milla et al., 1993). Nevertheless, for consistency, the conservative mutants were also recloned into the st11 tail background. Each of these mutants and wild-type Arc-st11 were purified to greater than 95% homogeneity by affinity and ion-exchange chromatography.

The oligomeric states of the mutants were determined by sedimentation equilibrium experiments. Wild-type Arc-st11 and the IV37, VI41, and IV37/VI41 mutants sediment as dimers at protein concentrations of 10 μM (Table 2). Each of these four proteins also had a similar fluorescence spectrum with a center of spectral mass near 336 nm for emission between 300 and 400 nm (Table 2). The wild-type and IV37 spectra are shown in Figure 2. Arc contains a single tryptophan residue (Trp14) which is buried in the native structure of the wild-type protein. The similarity of the mutant and wild-type fluorescence spectra suggests that the Trp14 side chains are also buried in the mutants. The Arc-st11, IV37, VI41, and IV37/VI41 proteins also have similar CD spectra (see Figure 3 for wild-type and IV37 spectra), suggesting that each protein has a similar secondary structure.

Unlike wild type and the conservative mutants, the IA37, VA41, and IA37/VA41 mutants sediment with molecular weights near those expected for monomers and have fluorescence emission spectra (center of mass near 350 nm) similar to those observed for pH-unfolded or urea-unfolded Arc (Table 2; Bowie & Sauer, 1989a; Milla & Sauer, 1994). Compared to the wild-type fluorescence spectrum, the spectra for the alanine mutants have decreased intensities and are red-shifted to positions that are expected if Trp14 is solvent exposed (see Figure 2 for the IA37 spectrum). The CD spectra of the alanine mutants show approximately 40% of the ellipticity at 222 nm of wild type or the other mutants, indicating significantly reduced α-helical contents (see Figure 3 for the IA37 spectrum). The residual secondary structure in the alanine mutants is greater than that observed for pH- or urea-denatured Arc. After pH or urea-denaturation, wild-type Arc has less than 10% of the ellipticity at 222 nm of native Arc (Bowie & Sauer, 1989c; Milla & Sauer, 1994).

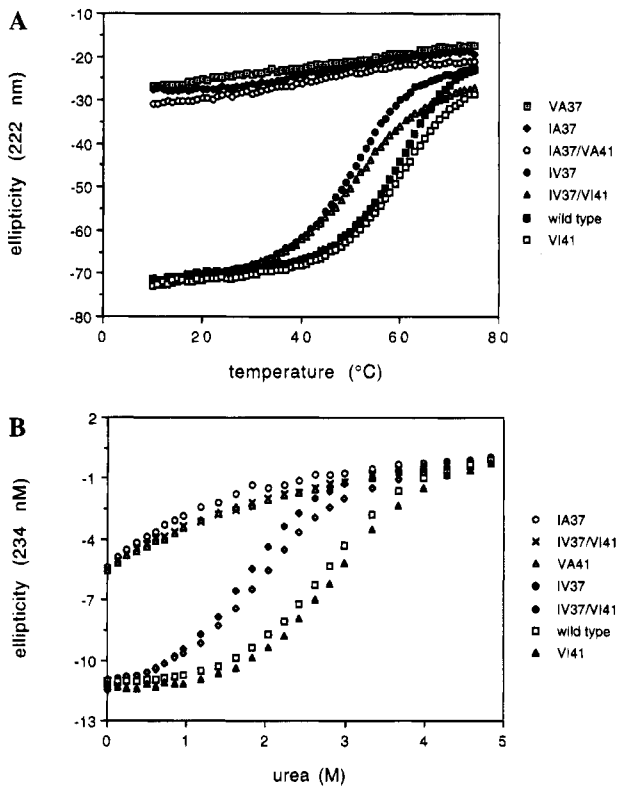

FIGURE 4: Equilibrium unfolding of wild-type Arc-st11 and mutants monitored by changes in circular dichroism. (A) Thermal denaturation (10 μM protein, 50 mM Tris-HCl (pH 7.5), 250 mM KCl, 0.2 mM EDTA); (B) Urea denaturation (10 μM protein, 25 °C, 50 mM Tris-HCl (pH 7.5), 250 mM KCl, 0.2 mM EDTA).

Effects of Conservative Substitutions on Arc Stability and Folding. Although the IV37, VI41, and IV37/VI41 mutant dimers have CD and fluorescence spectra similar to that of the wild-type Arc dimer, the mutations could affect protein stability. To test this possibility, equilibrium denaturation assays were performed. In thermal denaturation experiments, wild-type Arc-st11 and the VI41 mutant had t_m values within 1 °C, while the IV37 and IV37/VI41 mutants had t_m values reduced by roughly 10 °C (Figure 4A; Table 3). Calculation of ΔG_u and $\Delta\Delta G_u$ values at the wild-type t_m indicates that the IV37 and IV37/VI41 mutants are less stable than wild type by about 1.6 kcal/mol of dimer. In urea denaturation experiments, wild type and the VI41 mutant had similar C_m values, while the C_m values of the IV37 and IV37/VI41 mutants were reduced by approximately 1 M urea (Figure 4B). $\Delta\Delta G_u$ values calculated from the urea data are similar to those obtained in the thermal experiments and indicate that the IV37 and IV37/VI41 mutants are less stable than wild type by roughly 1.7−2.0 kcal/mol of dimer (Table 3). These data suggest that the loss of the δ-CH_3 methyl group that occurs in the IV37 single mutant is destabilizing, whereas addition of a new δ-CH_3 methyl group in the VI41 mutant is relatively neutral. The stability of the IV37/VI41 double mutant differs only slightly (by 0.07 and 0.36 kcal/mol of dimer in thermal and urea denaturation, respectively) from the stability calculated on the basis of additive effects of the single mutants. These small differences from additivity could arise because the 37 and 41 side chains interact (Horvitz & Fersht, 1990; Wells, 1990) but are probably not significant given the errors in the stability measurements.

Kinetic experiments were performed to determine the effects of the IV37, VI41, and IV37/VI41 mutations on the rates of protein folding and unfolding. In pH-jump refolding experiments, Arc-st11 and the VI41 mutant refold at comparable rates (Table 4), while the IV37 and IV37/VI41 mutants refold approximately 3-fold more slowly. In urea-jump unfolding experiments, the IV37, VI41, and IV37/VI41 mutants unfold roughly 7-fold, 3-fold, and 18-fold more rapidly than wild-type Arc-st11 (Figure 5; Table 4). Thus, the IV37 and IV37/VI41 mutations affect both the rate of protein folding and the rate of protein unfolding, while the VI41 mutation has a significant effect only on the rate of protein unfolding.

The Alanine-Substituted Mutants Exist in a Physiologically Denatured State. As mentioned above, the IA37, VA41, and IA37/VA41 mutants are monomeric at a 10 μM protein concentration, have reduced levels of secondary structure, and have fluorescence emission spectra that indicate that their tryptophan side chains are solvent exposed. Attempts to denature these mutants using heat (Figure 4A) or urea (Figure 4B) result in non-cooperative transitions. In both cases, the ellipticities of the alanine-substituted mutants approach that of wild type at high temperature or high urea concentration, where the wild-type protein is substantially denatured. All of these properties are consistent with the alanine mutants existing in a physiologically denatured state that is similar in many ways to the molten globule state (Kuwajima, 1989; Ptitsyn et al., 1990). ANS, a hydrophobic dye, is known to bind the molten globule conformations of many proteins in preference to the native protein (Rosen & Weber, 1969; Goto & Fink, 1989). As shown in Figure 6, the IA37, VA41, and VA37/VA41 Arc mutants bind ANS with apparent equilibrium dissociation constants of roughly 60 μM, but wild-type Arc in its native state also binds ANS with a similar apparent K_d.

Although the alanine mutants appear to exist in a monomeric denatured state at a protein concentration of 10 μM, it seemed possible that protein folding and dimerization might occur at higher protein concentrations. Indeed, sedimentation experiments performed at a protein concentration of 400 μM gave molecular weights intermediate between those expected for monomers and dimers for each of the alanine mutants (Table 2). Equilibrium constants and ΔG_u values for the concerted unfolding/dissociation reaction were calculated from the 400 μM sedimentation data and are listed in Table 2. From these values, we calculate that the expected mole fractions of dimer for the IA37, VA41, and IA37/VA41 mutants are 4%, 8%, and 3%, respectively, at 10 μM and 48%, 60%, and 40%, respectively, at 400 μM. Fluorescence emission spectra were also recorded for these mutants at protein concentrations of 400 μM (Table 2). The center of spectral mass for the VA41 mutant is blue shifted roughly 56% toward the value expected for wild type, an amount consistent with the estimate of 60% dimer. However, the IA37 and IA37/VA41 mutants are blue shifted only 12−22% toward the wild-type value, and yet they are estimated to be 40−48% dimeric under these conditions. This suggests that Trp14 may be more solvent exposed in the IA37 and IA37/VA41 mutant dimers than in the wild-type dimer. This could arise because truncation of the Ile37 side chain opens a solvent channel to Trp14 (the side chains of 14 and 37 are in contact in the wild-type structure), because the mutant dimers fold differently from wild type, or because the

Table 3: Stability Parameters (±SD) for Arc-st11 and Conservative Mutants Determined at 10 μM Protein Concentration, pH 7.5, and 250 mM KCl[a]

	thermal			urea		
protein	t_m (°C)	ΔH_u (kcal/mol)	$\Delta\Delta G_u$[a] (kcal/mol)	C_m (M)	m [(kcal/(mol M)]	$\Delta\Delta G_u$[b] (kcal/mol)
Arc-st11	57.5 (0.7)	56.5 (2.8)		2.56 (0.12)	1.64 (0.07)	
IV37	48.6 (0.4)	50.3 (2.3)	1.56 (0.32)	1.46 (0.07)	1.50 (0.07)	2.00 (0.23)
VI41	56.0 (0.8)	52.1 (2.9)	0.23 (0.36)	2.77 (0.16)	1.49 (0.08)	0.07 (0.26)
IV37/VI41	46.1 (0.5)	37.4 (2.3)	1.59 (0.28)	1.65 (0.08)	1.51 (0.07)	1.71 (0.24)

[a] Values of t_m and ΔH_u (at t_m) were determined from thermal melts; values of m (the slope of a plot of ΔG_u vs [urea]) and C_m (the urea concentration at 50% denaturation) were determined from urea unfolding experiments at 25 °C. $\Delta\Delta G_u^a = \Delta G_u$(protein) − ΔG_u(wild type) at the wild-type t_m. $\Delta\Delta G_u^b = \Delta G_u$(protein) − ΔG_u wild type) in the absence of urea at 25 °C

Table 4: Rate Constants (±SD) for Refolding (k_f) and Unfolding (k_u) at pH 7.5, 25 °C, and 250 mM KCl[a]

protein	k_f (M^{-1} s^{-1})	k_u (s^{-1})	m_u [kcal/(mol M)]
Arc-st11	8.4(1.8) × 10^6	0.11(0.01)	0.41(0.01)
IV37	2.7(0.5) × 10^6	0.74(0.07)	0.35(0.01)
VI41	9.6(0.8) × 10^6	0.37(0.02)	0.37(0.01)
IV37/VI41	2.6(0.7) × 10^6	1.97(0.34)	0.31(0.02)

[a] Values of k_f were determined from pH-jump experiments. Values of k_u were determined by extrapolation of the data shown in Figure 5 to 0 M urea. m_u is the slope of a plot of RT ln(k_u) vs [urea].

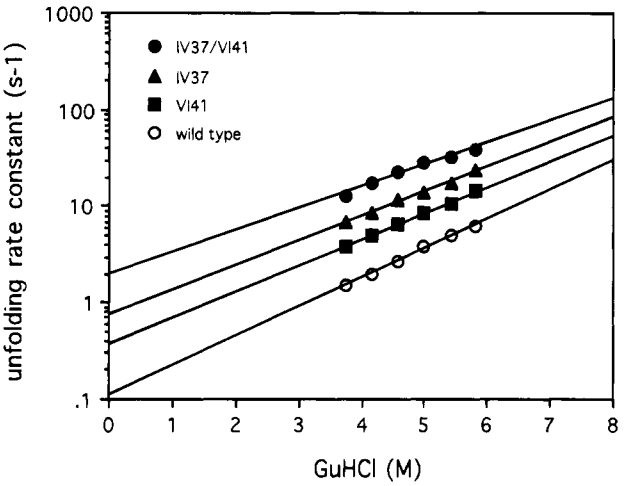

FIGURE 5: Urea dependence of the unfolding rate constants for wild-type Arc-st11 and mutants IV37, VI41, and IV37/VI41. Rate constants were measured at different urea concentrations at 25 °C, in buffer containing 50 mM Tris-HCl (pH 7.5), 250 mM KCl, and 0.2 mM EDTA.

mutants still have fluctuating structures, similar to those of molten globules, in the dimeric state.

DISCUSSION

To probe the importance of the Ile37 and Val41 packing interactions in Arc structure and stability, we have used combinatorial mutagenesis to test whether other combinations of hydrophobic residues were allowed at these positions. Of the 64 permutations permitted by the mutagenesis strategy, only the wild-type residue combination was recovered in a fully active protein. Seven additional combinations were found in mutants that were partially active and/or stable enough to be expressed at reasonable levels in *E. coli*. Each of these mutants had core volumes within 2 methylene equivalents of wild type. Only Val and Leu were allowed as replacements for Ile37, and only Ile, Leu, and Thr were allowed as replacements for Val41. The stability of Arc mutants *in vitro* is generally correlated with their steady-state level of expression *in vivo* because unstable mutants

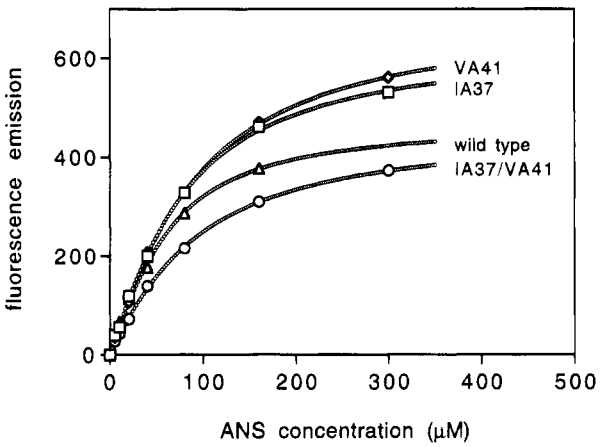

FIGURE 6: Binding of ANS to wild-type Arc-st11 and the IA37, VA41, and IA37/VA41 mutants monitored by the fluorescence emission (arbitrary units) of the ANS at 480 nm following excitation at 400 nm. Assays were performed using 50 μM protein, 50 mM Tris-HCl (pH 7.5), 250 mM KCl, and 0.2 mM EDTA, at 25 °C. The lines drawn in small symbols are theoretical curves calculated with K_d/maximal fluorescence values of 51 μM/640 (IA37), 62 μM/696 (VA41), 32 μM/477 (wild-type Arc-st11), and 62 μM/461 (IA37/VA41).

are subject to degradation (Vershon et al., 1986; Bowie & Sauer, 1989a,b). Since 41 of the 48 mutants that were isolated showed no detectable intracellular expression, it seems likely that most combinations of hydrophobic side chains at positions 37 and 41 result in unstable Arc proteins.

The results cited above suggest that the side chains of residues 37 and 41 need to maintain a near wild-type core volume and to have the correct shapes to fill this volume. The first conclusion is supported by measurements of the stability of the purified mutant proteins. The IV37, VA41, and IA37 mutations remove two, four, and six CH_2/CH_3 groups per dimer, respectively, and the corresponding mutant proteins have stabilities reduced by roughly 2.0, 6.0, and 6.4 kcal/mol of dimer compared with wild type. The IV37/VI41 mutant, in which the wild-type core volume is maintained but the positions of the Ile and Val side chains are switched, is also less stable than wild type by roughly 1.7 kcal/mol of dimer, supporting the conclusion that packing between residues 37 and 41 is also an important factor in determining the stability of Arc.

Are the Arc core mutations in which smaller side chains are introduced any more or less deleterious than comparable mutations in the hydrophobic cores of other proteins? To address this question, $\Delta\Delta G_u$ values for six Ile → Val, five Val → Ala, and five Ile → Ala mutations at completely buried positions in the phage f1 gene V dimer, staphylococcal nuclease, and barnase were taken from the data compiled in Table 5 of Serrano et al. (1992). To allow comparisons

between dimeric and monomeric proteins, the $\Delta\Delta G_u$ values for the gene V and Arc mutants were divided by 2, since each mutation in these proteins affects two side chains in the dimer. With these corrections, the destabilization caused by the Arc mutations can be seen to be very similar to those reported for the other proteins. The single Ile → Val and Ile → Ala changes at position 37 of Arc are destabilizing by 1.0 and 3.2 kcal/mol, respectively; these same changes in the other proteins cause average destabilizations (±SD) of 1.29 (±0.45) and 3.53 (±1.17) kcal/mol, respectively. Similarly, the single Val → Ala change at Arc position 41 reduces stability by 3.0 kcal/mol compared with an average reduction of 2.53(± 0.91) kcal/mol in gene V, barnase, and nuclease. Hence, the Arc mutants studied here have stabilities close to those expected for mutants bearing similar substitutions in other systems. Hydrophobic core mutations in which the side chain is truncated are thought to decrease stability both by reducing the amount of buried hydrophobic surface and by leaving cavities in the protein (Eriksson et al., 1992). Destabilization caused by changes in secondary structure propensity are also possible, but the mutant alanines in Arc would be expected to be more helix stabilizing than the wild-type β-branched residues which they replace at positions 37 and 41 (O'Neil & DeGrado, 1990; Blaber et al., 1993b). Thus, any changes in helical propensity should act to lessen the severity of the packing defects caused by these mutations.

Although the destabilization observed for the IV37, IA37, and VA41 mutants of Arc is not atypical, the overall packing restrictions for positions 37 and 41 appear more severe than in other proteins. For example, in the bacteriophage f1 gene V protein, two core residues, Val35 and Ile47, can be replaced with at least 27 different combinations of alanine, valine, methionine, leucine, isoleucine, and phenylalanine, and the protein maintains activity despite decreases in stability ranging from 0.6 to 7.1 kcal/mol of dimer (Sandberg & Terwilliger, 1991). In the N-terminal domain of λ repressor, randomization of three interacting core residues (Val36, Met40, and Val47) using valine, methionine, leucine, isoleucine, and phenylalanine resulted in at least 56 mutants with detectable biological activity (Lim & Sauer, 1991). At five interacting core positions in T4 lysozyme (Leu121, Ala129, Leu133, Val149, and Phe153) at least 106 different side-chain combinations are tolerated (Baldwin et al., 1993). The restrictions in the allowed packing interactions for residues 37 and 41 of Arc are reminiscent of those seen at a helix−helix dimerization interface in the N-terminal domain of phage λ repressor, where only the wild-type sequence (Ile84/Met87) and one conservative mutant (Ile84/Leu87) were recovered as active sequences following complete randomization of both positions (Reidhaar-Olson & Sauer, 1988).

Why do some residues involved in hydrophobic packing appear to be highly restricted, while others tolerate a wide variety of different packing combinations? Some of the differences undoubtedly depend on the inherent stability of the protein being studied, since a protein that is more stable can tolerate greater destabilization before denaturation ensues. Similarly, the stringency of the biological selection or screen that is used to define activity will influence the results, as more stringent selections will permit fewer sequences and *vice versa*. Nevertheless, it also seems likely that there are structural determinants of tolerance to substitution. The crystal structures of a packing variant of λ repressor (Lim et al., 1994) and several variants of T4 lysozyme (Baldwin et al., 1993) show that the changes in mutant side-chain packing are accommodated, in part, by movements of the polypeptide main chain. Perhaps restrictions in packing arise when the changes in main-chain conformation required to accommodate the mutant side chain become too energetically costly. It is also interesting that in both the λ repressor and Arc cases the packing involves symmetry-related side chains at a subunit interface. In these cases, it is probably more difficult to achieve efficient packing and also to satisfy the symmetry constraints. Another factor in tolerance toward hydrophobic substitutions may be the accessibility of alternative structures. In the GCN4 leucine-zipper dimer, hydrophobic substitutions are functionally tolerated at individual core positions at the coiled-coil interface (Hu et al., 1990), but proteins containing multiple, conservative substitutions at these positions change structure and form trimeric or tetrameric coiled-coils instead of dimers (Harbury et al., 1993).

In computer modeling studies, we were able to introduce the Val41 → Ile change into the wild-type structure of Arc with reasonable rotamer angles ($\chi_1 = \chi_2 = -60$, compared with $\chi_1 \approx 170$ for Val41) without introducing van der Waals contacts of less than 3.3 Å. Nevertheless, some structural changes are probably required to permit the additional δ-CH_3 groups to pack efficiently within the mutant core, because if they could be accommodated without strain, then the increased burial of hydrophobic surface in the mutant would have been expected to result in a protein that was more stable than wild type. In fact, the VI41 mutant is slightly less stable than wild type. Such structural changes could also explain why the VI41 mutant is not fully active and why it unfolds more rapidly than wild type.

The results reported here show that interactions mediated by the δ-CH_3 group of Ile37 are important in determining the stability of Arc (the IV37 mutant is approximately 2 kcal/mol of dimer less stable than wild type), in determining the energy difference between the native state and the transition state for unfolding (the IV37 mutant unfolds 7-fold faster than wild type), and in determining the energy difference between the unfolded state and the transition state for folding (the IV37 mutant refolds 3-fold slower than wild type). Because the IV37 mutation affects both the folding and unfolding rates, the δ-CH_3 group of Ile37 must be involved in interactions at or before the transition state and also in interactions after the transition state. Since the rate-determining step for Arc refolding involves dimerization (Milla & Sauer, 1994), it makes physical sense that Ile37, which forms part of the dimer interface in native Arc, should be involved in this critical step in folding. However, there is no reason to believe that the hydrophobic core is fully packed at this point during folding. In fact, the urea dependence of the Arc folding and unfolding rates suggests that approximately 75% of the hydrophobic surface is buried between the denatured state and the transition state, and 25% is buried between the transition state and the native state (Milla & Sauer, 1994). The kinetic results observed for the IV37 mutant are consistent with a model in which the δ-CH_3 group of Ile37 is involved in loose hydrophobic interactions which are formed in the transition state but also participates in the final tight packing of the core which occurs later in folding.

Arc mutants bearing the IA37, VA41, and IA37/VA41 substitutions are sufficiently destabilized that they exist almost exclusively as monomers at a protein concentration of 10 μM (the equilibrium constants for dimer unfolding/dissociation for these mutants are 0.2−0.7 mM; Table 2). This provides an opportunity to examine the properties of Arc in a monomeric state under physiological conditions. The IA37, VA41, and IA37/VA41 monomers exhibit many of the properties expected for molten globules (Kuwajima, 1989; Ptitsyn et al., 1990). These include the absence of cooperative transitions in urea- or thermal-denaturation experiments, the presence of reduced but significant amounts of α-helical secondary structure, and the absence of tertiary packing for Trp14. Fluorescence and ^{1}H-NMR studies of pressure-denatured Arc have also shown that these wild-type monomers have many properties expected of molten globules, including compact shape, loss of tertiary contacts, exposure of Trp14, and retention of some secondary structure (Silva et al., 1992; Peng et al., 1993). One difference between the Arc monomers and many molten globules concerns the binding of hydrophobic dyes. Although the mutant monomers described here bind ANS and the pressure-denatured wild-type monomers bind bis-ANS (Silva et al., 1992), the native Arc dimer also binds both of these hydrophobic dyes with similar affinities. From a structural perspective, it is not surprising that Arc monomers are denatured, since each monomer in the native dimer wraps around the other and a substantial portion of the hydrophobic core only becomes buried in the dimer. It is also not surprising that Arc monomers have many molten globule-like properties, as this is probably the default conformation of most denatured proteins under physiological conditions. At present, we do not know whether Arc monomers retain specific elements of the secondary or tertiary structure of the native protein or whether they are intermediates in the concerted dimerization/folding reaction.

ACKNOWLEDGMENT

We thank Bronwen Brown, Carey Waldburger, Alan Davidson, and Joel Schildbach for help and discussions; Chuck Kissenger, Upel Obeysekare, and Carl Pabo for providing the coordinates of the Arc dimer; and Paul Schimmel for use of equipment.

REFERENCES

Baldwin, E. P., Hajiseyedjavadi, O., Basse, W. A., & Matthews, B. W. (1993) *Science 262*, 1715−1718.

Becktel, W. J., & Schellman, J. A. (1987) *Biopolymers 26*, 1859−1877.

Bevington, P. R. (1969) *Data Reduction and Error Analysis for the Physical Sciences,* McGraw-Hill, New York.

Blaber, M., Lindstrom, J. D., Gassner, N., Xu, J.,Heinz, D. W., & Matthews, B. W. (1993a) *Biochemistry 32*, 11363−11373.

Blaber, M., Zhang, X. J., & Matthews, B. W. (1993b) *Science 260*, 1637−1640.

Bonvin, A. M., Vis, H., Breg, J. N., Burgering, M. J., Boelens, R., & Kaptein, R. (1994) *J. Mol. Biol. 236*, 328−341.

Bowie, J. U., & Sauer, R. T. (1989a) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2152−2156.

Bowie, J. U., & Sauer, R. T. (1989b) *J. Biol. Chem. 264*, 7596−7602.

Bowie, J. U., & Sauer, R. T. (1989c) *Biochemistry 28*, 7139−7143.

Breg, J. N., van Opheusden, J. H. J., Burgering, M. J., Boelens, R., & Kaptein, R. (1990) *Nature 346*, 586−589.

Brenstein, R. J. (1989) *NonLin for Macintosh*, Robelko Software, Carbondale, IL.

Brown, B. M., Bowie, J. U., & Sauer, R. T. (1990) *Biochemistry 29*, 11189−11195.

Brown, B. M., Milla, M. E., Smith, T. L., & Sauer, R. T. (1994) *Nature Struct. Biol. 1*, 164−168.

Eriksson, A. E., Baase, W. A., Zhang, X.-J., Heinz, D. W., Blaber, M., Baldwin, E. P., & Matthews, B. W. (1992) *Science 255*, 178−183.

Fersht, A. R., Matouschek, A., & Serrano, L. (1992) *J. Mol. Biol. 224*, 771−782.

Goto, Y., & Fink, A. L. (1989) *Biochemistry 28*, 945−952.

Harbury, P. B., Zhang, T., Kim, P. S., & Alber, T. (1993) *Science 262*, 1401−1407.

Horovitz, A., & Fersht, A. R. (1990) *J. Mol. Biol. 214*, 613−617.

Hu, J. C., O'Shea, E. K., Kim, P. S., & Sauer, R. T. (1990) *Science 250*, 1400−1403.

Johnson, M., & Frasier, S. (1985) *Methods Enzymol. 117*, 301−342.

Knight, K. L., Bowie, J. U., Vershon, A. K., Kelley, R. D., & Sauer, R. T. (1989) *J. Biol. Chem. 264*, 3639−3642.

Kuwajima, K. (1989) *Proteins 6*, 87−103.

Laue, T. M., Shah, B. D., Ridgeway, T. M., & Pelletier, S. L. (1992) in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S., Rowe, A., & Horton, J.,) Eds. pp 90−125, Royal Society of Chemistry, Cambridge.

Lim, W. A., & Sauer, R. T. (1991) *J. Mol. Biol. 219*, 359−376.

Lim, W. A., Hodel, A., Sauer, R. T., & Richards, F. M. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 423−427.

Matthews, B. W. (1993) *Annu. Rev. Biochem. 62*, 139−160.

Milla, M. E., & Sauer, R. T. (1994) *Biochemistry 33*, 1125−1133.

Milla, M. E., Brown, B. M., & Sauer, R. T. (1993) *Protein Sci. 2*, 2198−2205.

Milla, M. E., Brown, B. M., & Sauer, R. T. (1994) *Nature Struct. Biol. 1*, 518−523.

Mossing, M. C., Bowie, J. U., & Sauer, R. T. (1991) *Methods Enzymol. 208*, 604−619.

O'Neil, K. T., & DeGrado, W. F. (1990) *Science 250*, 646−651.

Peng, X., Jonas, J., & Silva, J. L. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 1776−1780.

Ptitsyn, O. B., Pain, R. H., Seminotnov, G. V., Zerovnik, E., & Razgulyaev, O. I. (1990) *FEBS Lett. 262*, 20−24.

Raumann, B. E., Rould, M. A., Pabo, C. O., & Sauer, R. T. (1994) *Nature 367*, 754−757.

Reidhaar-Olson, J. F., & Sauer, R. T. (1988) *Science 241*, 53−57.

Richards, F. M., & Lim, W. A. (1994) *Q. Rev. Biophys. 26*, 423−498.

Rosen, C.-G., & Weber, G. (1969) *Biochemistry 8*, 3915−3920.

Rosenberg, A. H., Lade, B. N., Chui, D.-s., Lin, S.-W., Dunn, J. J., & Studier, F. W. (1987) *Gene 56*, 125−135.

Sandberg, W. S., & Terwilliger, T. C. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1706−1710.

Serrano, L., Kellis, J. T., Cann, P., Matouschek, A., & Fersht, A. R. (1992) *J. Mol. Biol. 224*, 783−804.

Silva, J. L., Silveira, C. F., Correia, A., Jr., & Pontes, L. (1992) *J. Mol. Biol. 223*, 545−555.

Studier, F. W., & Moffatt, B. A. (1986) *J. Mol. Biol. 189*, 113−130.

Susskind, M. M., & Youderian, P. (1983) Bacteriophage P22 Antirepressor and Its Control, in *Lambda II* (Hendrix, R. W., Roberts, J. W., Stahl, F. W., & Weisberg, R. A., Eds.) pp 347−364, Cold Spring Harbor Laboratory Press, Cold.

Vershon, A. K., Youderian, P., Susskind, M. M., & Sauer, R. T. (1985) *J. Biol. Chem. 260*, 12124−12129.

Vershon, A. K., Bowie, J. U., Karplus, T. M., & Sauer, R. T. (1986) *Proteins 1*, 302−311.

Vershon, A. K., Liao, S-M., McClure, W. R., & Sauer, R. T. (1987) *J. Mol. Biol. 195*, 323−331.

Wells, J. A. (1990) *Biochemistry 29*, 8509−8517.

BI942210S