Evidence for Partial Secondary Structure Formation in the Transition State for Arc Repressor Refolding and Dimerization†

Alok K. Srivastava and Robert T. Sauer*

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

Received February 23, 2000; Revised Manuscript Received May 9, 2000

ABSTRACT: Structure formation and dimerization are concerted processes in the refolding of Arc repressor. The integrity of secondary structure in the transition state of Arc refolding has been investigated here by determining the changes in equilibrium stability and refolding/unfolding kinetics for a set of Ala → Gly mutations at residues that are solvent-exposed in the native Arc dimer. At some sites, reduced stability was caused primarily by faster unfolding, indicating that secondary structure at these positions is largely absent in the transition state. However, most of the Ala → Gly substitutions in the α-helices of Arc and a triple mutant in the β-sheet also resulted in decreased refolding rates, in some cases, accounting for the major fraction of thermodynamic destabilization. Overall, these results suggest that some regions of native secondary structure are present but incompletely formed in the transition state of Arc refolding and dimerization. Consolidation of this secondary structure, like close packing of the hydrophobic core, seems to occur later in the folding process. On average, Φ_F values for the Ala → Gly mutations were significantly larger than Φ_F values previously determined for alanine-substitution mutants, suggesting that backbone interactions in the transition state may be stronger than side chain interactions. Mutations causing significant reductions in the Arc refolding rate were found to cluster in the central turn of α-helix A and in the first two turns of α-helix B. In the Arc dimer, these elements pack together in a compact structure, which might serve as nucleus for further folding.

Studies of the way in which specific proteins fold and adopt their native tertiary and quaternary structure reveal an interesting variety of mechanisms (*1*−*5*). For example, some proteins display folding intermediates, whereas others fold in two-state reactions in which the denatured and native states are the only significantly populated species. Similarly, multimeric proteins may show distinct or concerted folding and oligomerization reactions. Even in cases in which intermediates cannot be studied directly, some information about the folding process can still be inferred from mutational analyses. Specifically, by analyzing how the equilibrium effects of mutations partition between the kinetic reactions for formation and loss of structure, it is often possible to probe whether certain interactions are absent, partially formed, or fully formed in the transition state (*6*, *7*). For example, some proteins displaying two-state folding still appear to fold in distinct steps because they have transition states in which some native interactions are nearly fully formed whereas others are almost completely absent.

The Arc repressor of bacteriophage P22 belongs to the ribbon−helix−helix family of DNA-binding proteins and folds as a homodimer in which the subunits intertwine and contribute equally to a single hydrophobic core (*8*−*10*). As a consequence, monomers cannot fold stably, and folding and dimerization are coincident processes (*11*, *12*). Each Arc subunit in the native dimer contains two α-helices (designated A and B) and a single β-strand, which pairs with its counterpart in the other subunit to form an antiparallel β-sheet (Figure 1). Arc displays two-state behavior in both equilibrium and kinetic studies of protein unfolding and refolding (*11*, *12*). The transition state is dimeric by the criterion that the Arc refolding rate is concentration-dependent and lies at a position along the reaction coordinate at which 70−80% of the final change in accessible surface has occurred (*12*). The equilibrium and kinetic consequences of a complete set of alanine-substitution mutations in Arc have been determined, revealing that destabilizing substitutions have roughly one-quarter of their energetic effect on the refolding rate and three-quarters of their effect on the unfolding rate (*13*, *14*). These studies suggest that most side chain interactions that stabilize the native Arc dimer are partially but weakly formed in the Arc transition state. The current view of the transition state for Arc refolding and dimerization is therefore one of a quasi-native structure in which core-packing and side chain hydrogen-bonding interactions are present but loosely or imperfectly formed.

In this paper, we test for formation of secondary structure during the refolding and dimerization of Arc repressor by using an Ala → Gly substitution strategy (*15*−*17*). In this method, the effects on stability and folding kinetics are determined for glycine versus alanine substitutions at solvent-exposed positions in the native structure. Because glycine can adopt many more backbone dihedral angles than alanine in the denatured state (*18*), Ala → Gly substitutions in α-helices and β-sheets should destabilize proteins by increas-

† Supported by NIH Grant AI-15706.

* To whom correspondence should be addressed at 68-571, 77 Massachusetts Avenue, M.I.T., Cambridge, MA 02139-4307; telephone: (617)-253-3163; fax: (617)-258-0673; e-mail: bobsauer@mit.edu.

10.1021/bi000423d CCC: $19.00
Published on Web 06/23/2000

ing the entropic cost of folding (*19*). Moreover, additional destabilization would be expected from the loss of packing and hydrophobic interactions mediated by the β-methyl group of alanine in the native structure. If a particular residue is unstructured in the transition state, then the destabilization caused by an Ala → Gly substitution at that position should affect the unfolding rate but not the folding rate and vice versa. Similarly, in the transition state, if the conformational flexibility of that residue has been partially reduced or the β-methyl group is engaged in some of its final packing interactions, then the glycine substitution would be expected to have effects on both the refolding and the unfolding rates. Studies of this type provide information about the protein backbone and therefore complement studies that examine the role of side chain interactions during the folding reaction.

MATERIALS AND METHODS

Construction of Mutants. Glycine-substituted Arc mutants were constructed by cassette mutagenesis in the *arc*-st11 gene of pSA700 (*20*, *21*). The identities of all site-directed mutants were verified by DNA sequencing. Arc variants were overexpressed in a *slyD*$^-$ derivative of *Escherichia coli* strain X90 [*F′ lacI*^{q} *lac*$^+$ *pro*$^+$*/ara* Δ*(lac-pro) nal1 argEam rif thi-1*], transformed with the appropriate mutant version of pSA700. Mutant proteins were purified to greater than 95% homogeneity by chromatography on nickel-NTA affinity columns as described (*20*). Protein concentrations in monomer equivalents per liter were calculated using an extinction coefficient at 280 nm of 6756 M^{-1} cm^{-1} (*22*).

Equilibrium Stability Measurements. Equilibrium denaturation reactions were assayed by changes in circular dichroism ellipticity and were at least 95% reversible. Urea denaturation (25 °C) and thermal denaturation were initially performed at a protein concentration of 10 μM in buffer containing 10 mM Tris-HCl (pH 7.5) and 250 mM KCl. For some unstable mutants, thermal denaturation experiments were also performed at higher protein concentrations, in which the native dimer is more highly populated. Denaturation data were fitted to equations for a two-state transition between denatured monomers and native dimers (*11*, *13*) using nonlinear least-squares fitting procedures.

Unfolding and Folding Kinetics. Kinetic experiments were performed at 25 °C using an Applied Photophysics DX17.MV stopped-flow instrument to monitor changes in tryptophan fluorescence at 320 nm (excitation 280 nm). For unfolding jumps, mutant proteins at concentrations of 10−60 μM in 10 mM Tris-HCl (pH 7.5) and 250 mM KCl were diluted into the same buffer containing 2−6 M urea. For refolding jumps, mutant proteins at concentrations of 10−60 μM in 4.8 M urea, 10 mM Tris-HCl (pH 7.5) and 250 mM KCl were diluted into the same buffer containing 0.8−4 M urea. Kinetic trajectories in refolding and unfolding experiments were fit to a two-state relaxation model involving a bimolecular refolding reaction and unimolecular unfolding reaction, using eqs 12 and 13 from Milla and Sauer (*13*). For each urea concentration used, the relaxation kinetics were constrained by the equilibrium constant determined experimentally for that urea concentration.

RESULTS

To probe the state of the polypeptide backbone in the transition state for Arc refolding and dimerization, we first selected surface positions in each element of secondary structure (Figure 1), where alanine substitutions had little or no effect on Arc stability (*13*, *14*). We then constructed single glycine-substitution mutations at each of these 15 positions and purified the corresponding mutant proteins for stability and kinetic studies. Two glycine substitutions were made in Arc's β-sheet (QG9 and RG13), seven glycine exchanges were made in α-helix A (RG16, EG17, DG20, RG23, KG24, EG26, and EG27), and six glycine mutations were made in α-helix B (NG34, SG35, QG39, KG46, KG47, and EG48) (Figure 1). The two β-sheet mutants were found to have stabilities similar to or greater than wild type. Hence, we also constructed and purified one triple-mutant protein in which each of the solvent-exposed residues in the β-sheet were mutated to glycine (QG9/NG11/RG13) and another triple-mutant protein in which these residues were changed to alanine (QA9/NA11/RA13). The stabilities of each of these mutants to urea denaturation were determined. In addition, the rates of protein unfolding and refolding were determined for most mutants. These equilibrium and kinetic data were then compared with the alanine-substitution data of Milla et al. (*13*, *14*) to determine the effects of Ala → Gly substitutions at each of the targeted sequence positions.

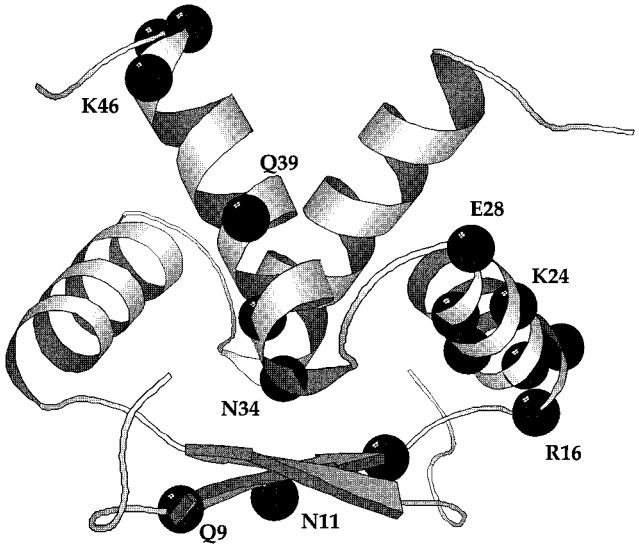

FIGURE 1: View of the Arc dimer as a polypeptide backbone ribbon trace. Sites at which glycine substitutions were introduced are highlighted as black spheres. This figure was prepared using the coordinates of an Arc dimer [1PAR.pdb; (*10*)] and MOLSCRIPT (*42*).

Effects of Glycine Mutations. Under native conditions, the glycine-substitution mutants had far-UV circular dichroism and fluorescence spectra that were indistinguishable from wild-type Arc (data not shown). Hence, these mutations do not appear to alter the native Arc structure in any significant fashion. The triple-glycine and triple-alanine β-sheet mutants also had spectroscopic properties equivalent to those of wild-type Arc.

The equilibrium denaturation of each mutant was assayed as a function of urea concentration at 25 °C, pH 7.5, and 250 mM KCl. These data were fit to obtain values for the equilibrium constant for unfolding and dimer dissociation (K_u) and the change in the free energy with denaturant ($m = \partial\Delta G_u/\partial$[urea]). Figure 2A shows several denaturation curves and the corresponding fits to the experimental data. Table 1 lists K_u and ΔG_u values for each of the mutant

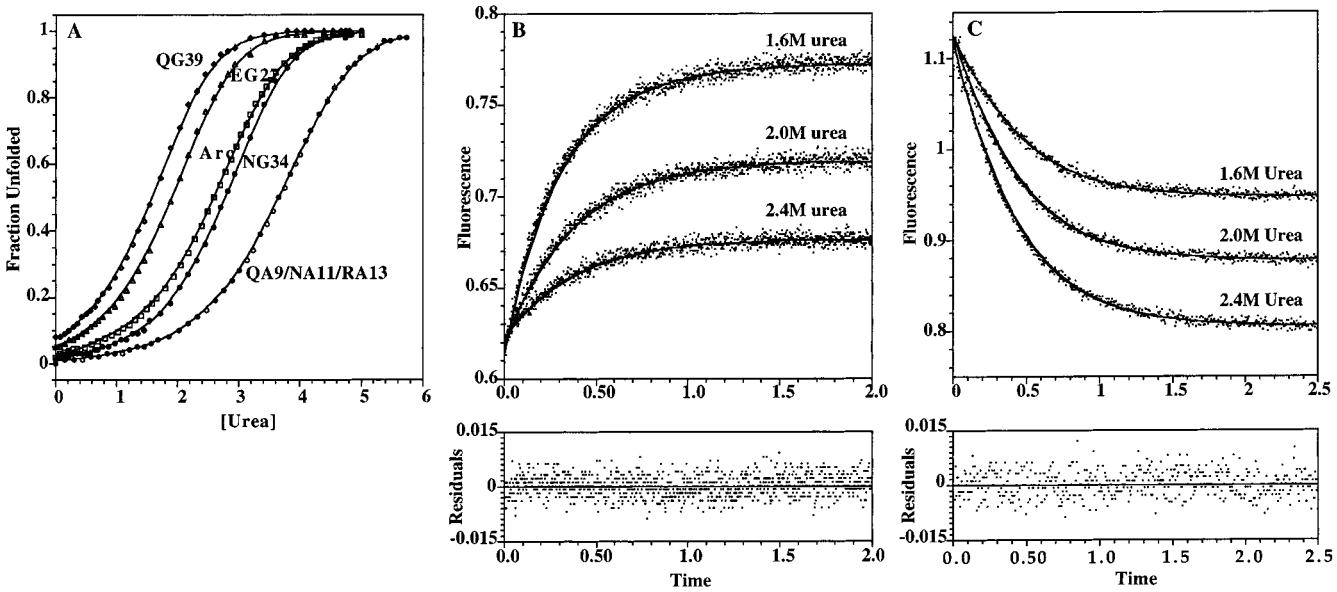

FIGURE 2: (A) Urea denaturation of Arc-st11 and selected mutant proteins. (B) Refolding kinetics of Arc-QG39-st11. (C) Unfolding kinetics of Arc-QG39-st11. In panels B and C, the solid lines are fits of the data to a relaxation reaction with bimolecular refolding and unimolecular unfolding components. The residuals of the fits for the 1.6 M urea data are shown below panels B and C.

Table 1: Properties of Glycine-Substitution Mutants[a]

mutant protein	K_u (M)	ΔG_u (kcal/mol)	$\Delta\Delta G_u^{wt-gly}$ (kcal/mol)	k_f (s^{-1} M^{-1})	k_u (s^{-1})	m (kcal/mol·M)	m_f (kcal/mol·M)	m_u (kcal/mol·M)	m_f/m
wild type	1.61E−08	10.62		1.27E+07	0.10	−1.48	−1.03	0.44	0.70
β-sheet									
QG9	5.13E−09	11.30	0.68	nd[b]	nd	−1.65	nd	nd	nd
RG13	8.86E−09	10.98	0.36	nd	nd	−1.35	nd	nd	nd
Q9A-N11A-R13A	2.10E−09	11.83	1.21	4.22E+07	0.04	−1.37	−1.11	0.31	0.81
Q9G-N11G-R13G	1.15E−08	10.82	0.20	2.18E+07	0.12	−1.38	−1.06	0.26	0.77
α-helix A									
RG16	2.23E−07	9.07	−1.55	3.76E+06	0.42	−1.24	−0.90	0.33	0.73
EG17	1.21E−07	9.43	−1.19	3.91E+06	0.24	−1.46	−0.92	0.54	0.63
DG20	1.80E−06	7.83	−2.79	7.96E+05	0.73	−1.46	−0.97	0.50	0.66
RG23	6.01E−07	8.48	−2.14	3.53E+06	1.06	−1.48	−1.06	0.41	0.72
KG24	1.02E−06	8.17	−2.45	5.10E+06	2.58	−1.67	−1.26	0.40	0.76
EG27	5.75E−08	9.87	−0.75	1.03E+07	0.30	−1.61	−1.03	0.57	0.64
EG28	2.81E−07	8.93	−1.69	9.76E+06	1.37	−1.54	−1.07	0.46	0.70
α−helix B									
NG34	5.49E−09	11.26	0.64	2.11E+07	0.06	−1.60	−1.09	0.50	0.68
SG35	3.52E−08	10.16	−0.46	7.30E+06	0.13	−1.51	−1.01	0.50	0.67
QG39	1.27E−07	9.40	−1.22	3.45E+06	0.22	−1.64	−1.11	0.53	0.67
KG46	2.41E−07	9.02	−1.60	1.01E+07	1.22	−1.36	−1.00	0.36	0.73
KG47	1.15E−07	9.46	−1.16	1.36E+07	0.78	−1.50	−1.08	0.42	0.72
EG48	1.29E−06	8.03	−2.59	8.24E+06	5.30	−1.09	−0.86	0.22	0.79

[a] pH 7.5, 250 mM KCl, 25 °C. [b] nd, not determined.

proteins and for wild-type Arc in the absence of urea. K_u and ΔG_u values for each mutant were also calculated at 50 °C from thermal denaturation experiments (data not shown) and were reasonably well-correlated with the values obtained at 25 °C from urea denaturation ($r = 0.97$ for ΔG_u). Inspection of the data in Table 1 shows that most of the glycine-substitution mutants destabilize the Arc dimer by roughly 0.5−3 kcal/mol-dimer, although three single mutants and the two triple mutants are actually more stable than wild-type Arc under these conditions.

Refolding and unfolding urea-jump experiments at final denaturant concentrations ranging from 0.8 to 4.5 M were used to assay the kinetic behavior of these mutants. These kinetic data were fit as relaxations to equilibrium involving a bimolecular refolding reaction and an unimolecular unfolding reaction (see Materials and Methods) to obtain values for the rate constants for refolding (k_f) and unfolding (k_u). Figure 2, panels B and C show a refolding and unfolding experiment, respectively, for the QG39 mutant and the quality of the theoretical fits. In all cases, the kinetics appeared two-state by the criteria that the amplitude of the kinetic transient was that expected from the equilibrium experiments, and the data fit well to a two-state relaxation model constrained by the equilibrium constant for unfolding and dimerization. Figure 3 shows $RT \ln(k_u)$ and $RT \ln(k_f)$ versus [urea] plots for five of the mutants and wild-type Arc. In general, these plots are reasonably linear ($r > 0.98$), although some curvature is evident in several of the data sets. Table 1 lists values of k_f and k_u for each mutant, obtained by extrapolation to 0 M denaturant. Most of the glycine-substitution mutations result in changes in both the refolding and the unfolding rates in comparison with wild-type Arc.

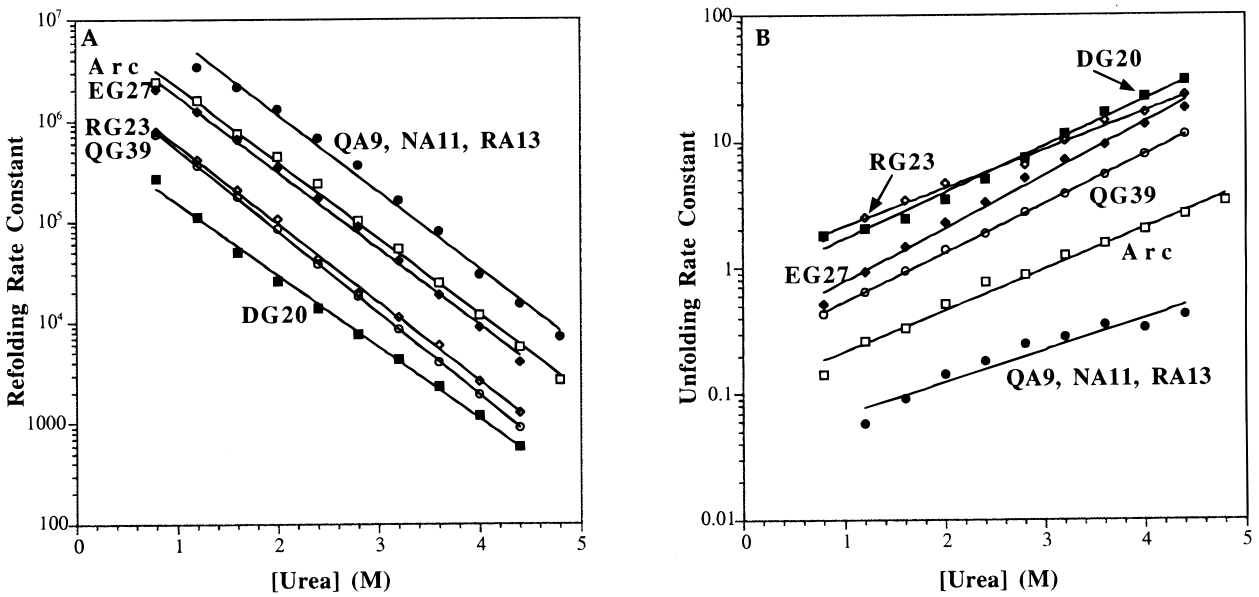

FIGURE 3: Urea dependence of the rate constants for refolding (A) and unfolding (B) of wild type and selected mutants. Refolding and unfolding experiments were performed at a protein concentration of 10 μM, 25 °C, pH 7.5, and 250 mM KCl.

The *m* values for equilibrium denaturation (*m*), for the kinetic refolding reaction (m_f), and for the kinetic unfolding reaction (m_u) are listed for each mutant in Table 1. In each case, these *m* values represent the slope of an *RT* ln(K_u) or *RT* ln(rate constant) vs urea plot. Each *m* value is an approximate measure of the change in solvent-accessible protein surface between the two states that determine the energetics of the reaction (*23*, *24*). For example, m_f reflects the change in surface burial between the denatured state and the transition state, and *m* depends on the difference in surface burial between the denatured state and the native state. The ratio m_f/m is a fractional measure of the surface buried in the transition state as compared to that buried in the native protein. A value of m_f/m close to 1 indicates similar solvent accessibilities for the transition-state ensemble and native protein, whereas a value close to 0 indicates that the transition state and the denatured protein have similar amounts of buried surface. For wild-type Arc, m_f/m was 0.7, suggesting that 70% of the surface area buried in native Arc is also solvent-inaccessible in the transition state. All of the glycine-substitution mutants had m_f/m values between 0.63 and 0.81 (Table 1). As a result, it appears that a substantial amount of surface area is buried in the transition state of each mutant. Moreover, the transition states of wild-type Arc and the mutants lie at roughly similar positions along a reaction coordinate defined by solvent-accessible surface area.

Effects of Ala → Gly Substitutions. To determine the effects of Ala → Gly substitutions, we calculated $\Delta\Delta G_u$ values ($\Delta G_u^{ala} - \Delta G_u^{gly}$) using the glycine-substitution data shown in Table 1 and the alanine-substitution data published by Milla et al. (*13*). For these comparisons, a reference concentration of 3 M urea (25 °C, pH 7.5, and 250 mM KCl) was chosen. Using this denaturant concentration minimizes errors in ΔG_u and $\Delta\Delta G_u$ values, by eliminating or reducing the length of extrapolations, and also allows comparison of refolding and unfolding kinetics at a denaturant concentration at which the rates of both refolding and unfolding are significant. Table 2 lists $\Delta\Delta G_u$ values, the relative rates of refolding of each alanine mutant and the corresponding glycine mutant ($k_{f-rel} = k_f^{ala}/k_f^{gly}$), the relative rates of unfolding of the alanine and glycine mutants ($k_{u-rel} = k_u^{gly}/k_u^{ala}$), and values representing the fractional energetic contribution of the relative refolding rates to the Ala → Gly equilibrium effect [$\Phi_F = RT \ln(k_{f-rel})/\Delta\Delta G_u$].

Table 2: Relative Stabilities and Folding Kinetics of Alanine and Glycine Substituted Mutants[a]

position	$\Delta\Delta G_u^{gly-ala}$ (kcal/mol)	k_f^{ala}/k_f^{gly}	k_u^{gly}/k_u^{ala}	Φ_F
β-sheet				
Q9-N11-R13	−1.02	2.55	2.21	0.54
α-helix A				
R16	−1.03	2.06	2.78	0.41
E17	−0.70	0.88	3.73	−0.11
D20	−1.94	4.39	6.10	0.45
R23	−1.59	2.39	6.21	0.32
K24	−2.49	4.13	11.27	0.34
E27	−1.22	0.96	8.23	−0.02
E28	−1.54	2.97	4.58	0.42
α-helix B				
N34	−0.49	3.83	0.60	1.62
S35	−1.57	6.61	2.17	0.71
Q39	−2.51	14.58	4.83	0.63
K46	−1.15	1.88	3.74	0.32
K47	0.25	0.67	0.98	
E48	0.09	1.01	0.85	

[a] pH 7.5, 25 °C, 250 mM KCl, and 3.0 M urea.

Inspection of Table 2 shows that 11 of the single glycine substitutions destabilize Arc by roughly 0.5−2.5 kcal/mol-dimer as compared to the alanine mutations. Increases of more than 2-fold in the unfolding rate constant are observed for 10 of these substitutions, and decreases of more than 2-fold in the second-order rate constant for refolding are seen for eight of these substitutions. The Φ_F values for these 11 Ala → Gly substitutions are quite heterogeneous, ranging from values close to 0 to values greater than 1. Significant changes in equilibrium stability, refolding rate, and unfolding rate are also observed for the triple glycine substitution in the β-sheet. At two residue positions at the C-terminal end of α-helix B, the single glycine mutants had approximately

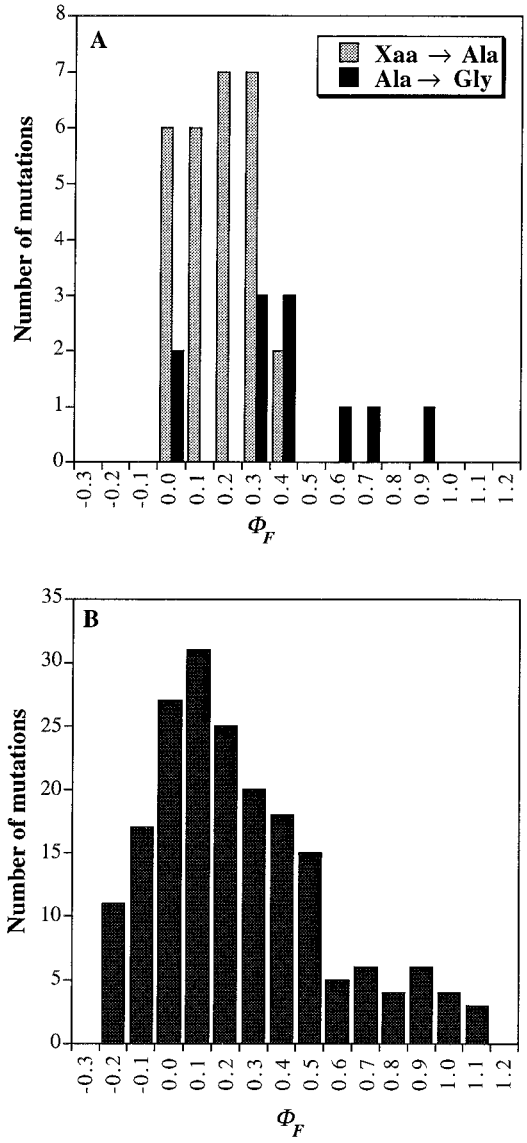

FIGURE 4: (A) Distribution of Φ_F values for Xaa → Ala (*14*) and Ala → Gly Arc mutants for which $\Delta\Delta G_u > 0.5$ kcal/mol-dimer. (B) Distribution of Φ_F values for 192 mutants in small monomeric proteins (*5*). Histograms show the total number of sites at which amino acid replacements give rise to Φ_F values in the indicated range.

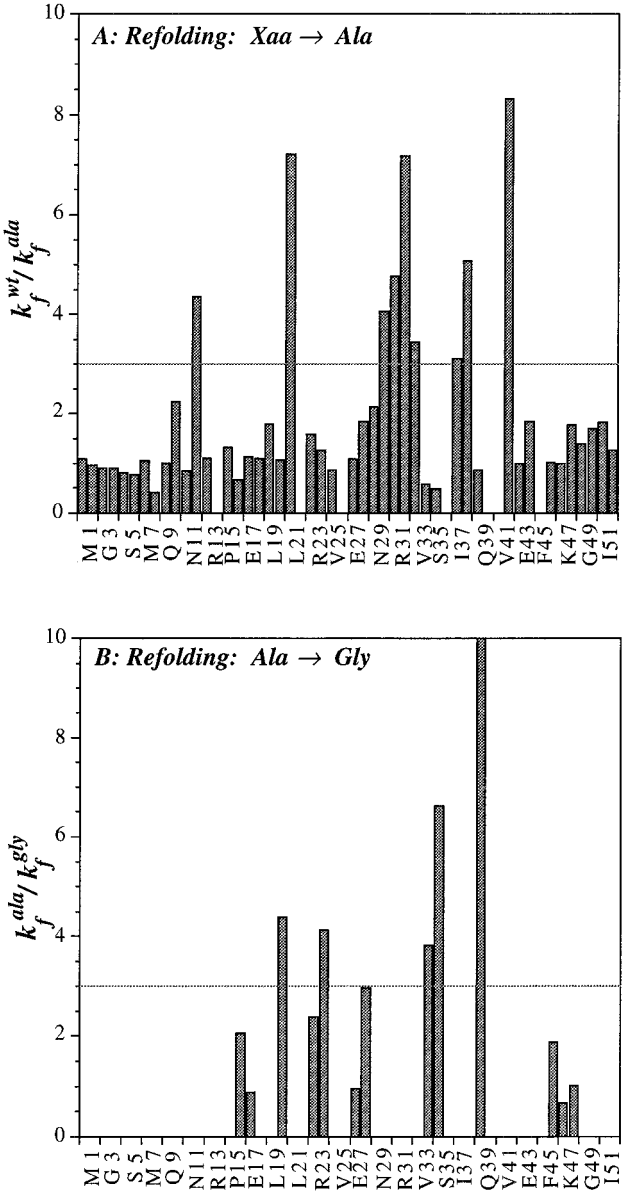

FIGURE 5: Effects on the relative rates of Arc refolding and dimerization of Xaa → Ala substitutions (A) and Ala → Gly substitutions (B).

the same stability as the corresponding alanine mutants and showed similar refolding and unfolding kinetics.

DISCUSSION

In the studies presented here, we have used Ala → Gly substitutions at surface positions in the Arc repressor dimer to probe the structure of the polypeptide backbone during the concerted refolding and dimerization reactions of this protein. In the α-helical regions of Arc, Ala → Gly substitutions were found to have a wide range of effects on equilibrium stability, ranging from a small stabilizing effect (−0.25 kcal/mol-dimer at residue 47) to a substantial destabilizing effect (2.5 kcal/mol-dimer at residue 24). Because each substitution is present twice in the Arc dimer, the average stability change per substitution is half of the dimer value. For the 13 Ala → Gly mutations in α-helical regions of Arc, the average stability change per substitution was 0.61 ± 0.43 kcal/mol. In other proteins, a group of 29 Ala → Gly substitutions on the surfaces of α-helices caused an average destabilization of 0.76 ± 0.44 kcal/mol (*16*, *17*, *25*−*27*). This average value is close to the destabilization of 0.73 ± 0.06 kcal/mol that is predicted solely from the loss in conformational entropy at 25 °C due to the increased flexibility of glycine as compared to alanine in the denatured state (*19*). In the two-stranded β-sheet of Arc, Ala → Gly replacements caused relatively small stability effects. On the basis of the triple mutants, the average destabilization per Ala → Gly substitution was only 0.16 kcal/mol. This value is significantly less than the average of 1.34 ± 0.68 kcal/mol observed for eight examples of Ala → Gly substitutions in β-sheets in other proteins (*28*−*30*). This discrepancy may be related to the fact that Arc's β-sheet has only two strands and the mutated residues are "edge" positions (*29*) or, as noted below, that this region has high native flexibility.

The wide variation in the stability effects caused by different Ala → Gly substitutions in Arc must, in principle, reflect differential changes in backbone flexibility, packing

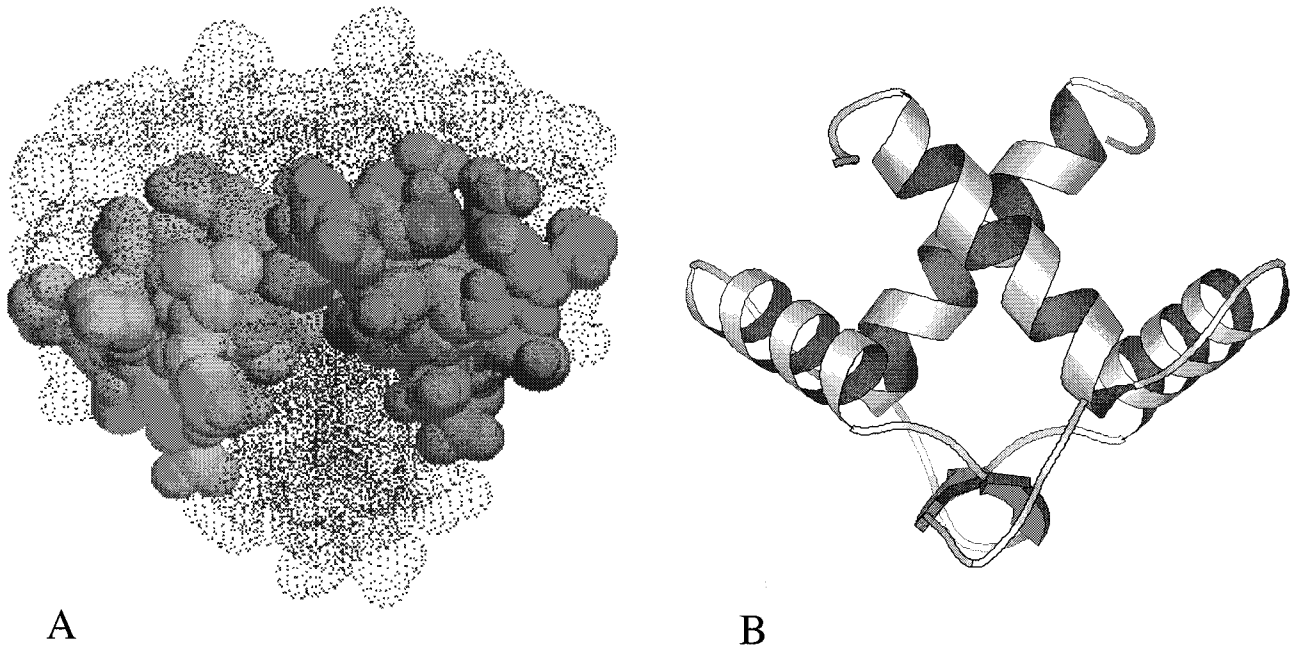

FIGURE 6: (A) View of the Arc dimer. Residues at which mutations cause the largest reductions in refolding rates (positions 20−24 and 30−42) are shown in space-filling representation with different shading for each subunit. Remaining residues are shown as a dot surface. (B) Arc dimer in the same orientation as in panel A but in a ribbon representation.

of the β-methylene group of alanine, and/or changes in solvation or structure between the native and denatured states of the alanine-substituted and glycine-substituted proteins. Attempts to rationalize the observed stability differences for individual Ala → Gly substitutions are speculative at best. Nevertheless, we note that Ala → Gly substitutions in the β-sheet of Arc and at the C-terminal end of helix B had the smallest stability effects and that these residues of Arc show highly dynamic behavior in NMR ^{15}N-relaxation and hydrogen−deuterium exchange studies (*31*). This correlation suggests that high levels of local backbone flexibility in native Arc may help to reduce the destabilizing effects of Ala → Gly substitutions in these regions.

Although Arc folds in an apparent two-state reaction, there is evidence for stepwise folding involving unstable intermediates (*12*, *32*−*36*). The heterogeneity of Φ_F values for the Ala → Gly mutants studied here also lends support to a stepwise folding mechanism. For example, the Ala → Gly substitutions at residues 17 and 27 increased the unfolding rate and had almost no effect on the refolding rate ($\Phi_F \approx 0$). This result indicates that these backbone positions, which are close to the beginning and the end of helix A, are unstructured in the transition state. At the opposite extreme, the substitution at residue 34 actually decreased the unfolding rate slightly and caused destabilization exclusively by slowing the refolding rate ($\Phi_F > 1$). This backbone position, which accepts an N-cap hydrogen bond from Ser32 in native Arc, appears to be completely structured in the transition state. By contrast, most of the other Ala → Gly mutants had substantial effects on both the refolding and unfolding rates. For instance, the triple Ala → Gly mutant in the β-sheet and the single mutants at positions 16, 20, 23, 24, 28, 35, 39, and 46 had an average Φ_F value of 0.46 ± 0.14. At these positions in Arc, the polypeptide backbone seems to be partially structured in the transition state for refolding. Alternatively, the folding reaction could involve parallel pathways with distinct transition states, and the intermediate Φ_F values could arise because a given interaction is made before the transition state in some pathways and after the transition state in other pathways (*37*, *38*). As shown in Figure 4A, the Φ_F values for the majority of the Ala → Gly mutants examined here are significantly higher than for the Arc alanine-substitution mutants. This result implies that backbone interactions, on average, are closer to completion or more highly consolidated than side chain interactions in the transition state of Arc. Comparison of Figure 4, panels A and B reveals that the overall distribution of Φ_F values for the Arc mutations is reasonably similar to the distribution observed for mutations in monomeric proteins (*5*). Hence, although refolding and dimerization are concerted processes for Arc, the general features of these reactions resemble simpler protein folding reactions.

Figure 5 shows the effects on relative refolding kinetics, measured here for the single Ala → Gly substitutions and previously for the Xaa → Ala (above) substitutions of Arc (*14*). By considering the sites at which mutations cause the largest decreases in refolding rates, we can begin to compose a rough sketch of the interactions that are important relatively early in the refolding and dimerization reaction. Backbone or side chain mutations at positions 12, 20, 21, 24, 30, 31, 32, 33, 34, 35, 38, 39, and 42 cause the largest decreases in the refolding rate. Residues 20−24 and 30−42 include all but one of these positions and also encompass positions 22, 36, 37, 41, and 42 where alanine-substitution mutants were too destabilizing to allow determination of refolding kinetics (*13*). As a result, it seems likely that the central turn of helix A (residues 20−24), the turn between the two helices (residues 30−34), and the first and second turns of helix B (residues 35−42) have already begun to form and to make native-like interactions in the transition state for Arc refolding and dimerization. Figure 6 shows that these regions of Arc form a reasonably compact substructure with substantial packing between the two subunits. Assembly of this substructure, relatively early during folding/dimerization, could

provide a foundation or nucleus for later folding events.

Plaxco et al. (*39*) have presented evidence that the average distance in sequence between residues that contact each other in the native structure may play a significant role in determining the rates of refolding of small, monomeric proteins. This makes sense if local topological interactions are the easiest to form in an entropic sense. In this regard, it is interesting that the residues in Arc, which contribute most to the refolding rate, fall in a reasonably small region (residues 20−42) comprising less than half of the protein. It is important to recall, however, that Arc folds as a dimer. Hence, many of the residues that interact both in the transition state and in the native dimer are in different subunits and are topologically unlinked. It is possible that the Φ_F values for the Ala → Gly backbone mutations are larger, on average, than for the alanine-substitution mutations, because the backbone mutations affect local interactions in the same chain, whereas many of the side chain mutations affect interactions between chains. Although Arc refolding requires the collision of two polypeptide chains, the folding and dimerization reaction at a protein concentration of 10 μM is as rapid as the folding of many monomeric proteins (*12*). Moreover, some mutants of Arc have refolding rates as much as 40-fold faster than wild type (*34*, *40*), and some single-chain variants of Arc have refolding rates that extrapolate into the microsecond time regime [(*41*); unpublished data]. Clearly, Arc folds and dimerizes in a reaction in which the proper backbone and side chain interactions are formed quite efficiently.

ACKNOWLEDGMENT

We thank Peter Chivers, Shari Spector, and Geeta Narlikar for comments on the manuscript; David Goldenberg for providing the compiled data shown in Figure 4B; and members of the lab for helpful discussions and advice.

REFERENCES

1. Baker, D., and Agard, D. A. (1994) *Biochemistry 33*, 7505−7509.
2. Fersht, A. R. (1997) *Curr. Opin. Struct. Biol. 7*, 3−9.
3. Baldwin, R. L., and Rose, G. D. (1999) *Trends Biochem. Sci. 24*, 26−33.
4. Baldwin, R. L., and Rose, G. D. (1999) *Trends Biochem. Sci. 24*, 77−83.
5. Goldenberg, D. P. (1999) *Nat. Struct. Biol. 6*, 987−990.
6. Goldenberg, D. P., Frieden, R. W., Haack, J. A., and Morrison, T. B. (1989) *Nature 338*, 127−132.
7. Matouschek, A., Kellis, J. T., Jr., Serrano, L., and Fersht, A. R. (1989) *Nature 340*, 122−126.
8. Breg, J. N., van Opheusden, J. H., Burgering, M. J., Boelens, R., and Kaptein, R. (1990) *Nature 346*, 586−589.
9. Bonvin, A. M., Vis, H., Breg, J. N., Burgering, M. J., Boelens, R., and Kaptein, R. (1994) *J. Mol. Biol. 236*, 328−341.
10. Raumann, B. E., Rould, M. A., Pabo, C. O., and Sauer, R. T. (1994) *Nature 367*, 754−757.
11. Bowie, J. U., and Sauer, R. T. (1989) *Biochemistry 28*, 7139−7143.
12. Milla, M. E., and Sauer, R. T. (1994) *Biochemistry 33*, 1125−1133.
13. Milla, M. E., Brown, B. M., and Sauer, R. T. (1994) *Nat. Struct. Biol. 1*, 518−523.
14. Milla, M. E., Brown, B. M., Waldburger, C. D., and Sauer, R. T. (1995) *Biochemistry 34*, 13914−13919.
15. Matthews, J. M., and Fersht, A. R. (1995) *Biochemistry 34*, 6805−6814.
16. Sosnick, T. R., Jackson, S., Wilk, R. R., Englander, S. W., and DeGrado, W. F. (1996) *Proteins 24*, 427−432.
17. Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L., and Oas, T. G. (1997) *Nat. Struct. Biol. 4*, 305−310.
18. Ramachandran, G. N., Ramakrishnan, C., and Sasikharan, V. (1963) *J. Mol. Biol. 7*, 95−99.
19. D'Aquino, J. A., Gomez, J., Hilser, V. J., Lee, K. H., Amzel, L. M., and Freire, E. (1996) *Proteins 25*, 143−156.
20. Milla, M. E., Brown, B. M., and Sauer, R. T. (1993) *Protein Sci. 2*, 2198−2205.
21. Brown, B. M., Milla, M. E., Smith, T. L., and Sauer, R. T. (1994) *Nat. Struct. Biol. 1*, 164−168.
22. Brown, B. M., Bowie, J. U., and Sauer, R. T. (1990) *Biochemistry 29*, 11189−11195.
23. Tanford, C. (1970) *Adv. Protein Chem. 24*, 1−95.
24. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) *Protein Sci. 4*, 2138−2148.
25. Hecht, M. H., Sturtevant, J. M., and Sauer, R. T. (1986) *Proteins 1*, 43−46.
26. Serrano, L., Neira, J. L., Sancho, J., and Fersht, A. R. (1992) *Nature 356*, 453−455.
27. Serrano, L., Sancho, J., Hirshberg, M., and Fersht, A. R. (1992) *J. Mol. Biol. 227*, 544−559.
28. Kim, C. A., and Berg, J. M. (1993) *Nature 362*, 267−270.
29. Smith, C. K., and Regan, L. (1995) *Science 270*, 980−982.
30. Otzen, D. E., and Fersht, A. R. (1995) *Biochemistry 34*, 5718−5724.
31. Nooren, I. M., Rietveld, A. W., Melacini, G., Sauer, R. T., Kaptein, R., and Boelens, R. (1999) *Biochemistry 38*, 6035−6042.
32. Peng, X., Jonas, J., and Silva, J. L. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 1776−1780.
33. Burgering, M. J., Hald, M., Boelens, R., Breg, J. N., and Kaptein, R. (1995) *Biopolymers 35*, 217−226.
34. Waldburger, C. D., Jonsson, T., and Sauer, R. T. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 2629−2634.
35. Jonsson, T., Waldburger, C. D., and Sauer, R. T. (1996) *Biochemistry 35*, 4795−4802.
36. Robinson, C. R., Rentzeperis, D., Silva, J. L., and Sauer, R. T. (1997) *J. Mol. Biol. 273*, 692−700.
37. Fersht, A. R., Itzhaki, L. S., elMasry, N. F., Matthews, J. M., and Otzen, D. E. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10426−10429.
38. Moran, L. B., Schneider, J. P., Kentsis, A., Reddy, G. A., and Sosnick, T. R. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 10699−10704.
39. Plaxco, K. W., Simons, K. T., and Baker, D. (1998) *J. Mol. Biol. 277*, 985−994.
40. Brown, B. M., and Sauer, R. T. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 1983−1988.
41. Robinson, C. R., and Sauer, R. T. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 5929−5934.
42. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

BI000423D