Articles

Equilibrium Stability and Sub-Millisecond Refolding of a Designed Single-Chain Arc Repressor†

Clifford R. Robinson and Robert T. Sauer*

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307

Received June 10, 1996; Revised Manuscript Received August 19, 1996⊗

ABSTRACT: Arc-L1-Arc is a single-chain variant of bacteriophage P22 Arc repressor in which a 15-residue linker joins the C-terminus of one subunit to the N-terminus of an otherwise identical subunit. Spectroscopic probes indicate that the native and denatured states of the single-chain protein are similar to those of the unlinked Arc dimer. In equilibrium experiments, Arc-L1-Arc denatures in a reaction without populated intermediate states as judged by the fits of the denaturation isotherms to a two-state model and by the coincidence of denaturation curves monitored by fluorescence and circular dichroism. Comparison of the equilibrium stabilities of Arc-L1-Arc and unlinked Arc gives an effective concentration of subunits in the denatured single-chain variant of 2.7 (± 0.7) mM. The kinetic refolding and unfolding reactions of Arc-L1-Arc also appear to proceed without populated intermediates. The rate constant for Arc-L1-Arc unfolding is about 2-fold faster than that of unlinked Arc, indicating that the linker mediates no significant contacts in the native structure that need to be broken to allow unfolding. As expected, the major effect of the linker occurs during the refolding reaction, where the effective subunit concentration calculated from the bimolecular and unimolecular refolding rate constants is 4.5 (± 1.8) mM. The transition states for the unfolding and refolding reactions of Arc-L1-Arc and wild-type Arc have similar solvent exposures as measured by the urea dependencies of the equilibrium and rate constants. In the absence of urea, the single-chain protein refolds very rapidly ($k_f \approx 10^4$ s^{-1}) in a reaction that is essentially complete in the sub-millisecond time regime.

For certain oligomeric proteins, the processes of folding and subunit association are tightly coupled and thus the thermodynamic stabilities and the folding rates of these proteins are concentration dependent. Well-studied examples include the coiled-coil domains found in many transcription factors, the gene V dimer of phage fd, the four-helix-bundle protein ROP, and the Arc repressor dimer of phage P22 (Bowie & Sauer, 1989; O'Shea et al., 1989, 1992; Liang & Terwillinger, 1991; Steif et al., 1993; Milla & Sauer, 1994). In each of these proteins, the three-dimensional structure shows that individual subunits wrap around each other or are intertwined in some fashion, with hydrophobic cores that are formed by side chains from multiple subunits (O'Shea et al., 1991; Steif et al., 1993; Raumann et al., 1994; Skinner et al., 1994). Dissociation of the oligomeric forms of these proteins results in unfolding, because the monomers are unable to form stable structures by themselves. In such cases, covalent linkage of the subunits by peptide linkers or disulfide bonds provides a strategy for increasing protein stability (O'Shea et al., 1989; Liang et al., 1993; Predki & Regan, 1995).

Arc-L1-Arc is a single-chain variant of P22 Arc repressor that was engineered by designing a gene with tandem copies of the coding sequence for the Arc monomer (53 residues) connected by DNA encoding a 15-residue, glycine-rich linker (Robinson & Sauer, 1996). Figure 1 shows one model for a linker conformation that could connect the C-terminus of the first subunit to the N-terminus of the second. The Arc-L1-Arc protein was found to be soluble, monomeric, and fully active in operator-binding and repression assays (Robinson & Sauer, 1996). In this paper, we compare the spectroscopic properties, equilibrium stabilities, and refolding and unfolding kinetics of Arc-L1-Arc with those of an unlinked Arc variant. These experiments suggest that the linker plays a passive, largely entropic role by keeping the denatured subunits at high local concentrations during folding. Comparison of the equilibrium stabilities of the linked and unlinked proteins gives an effective concentration of 2.7 mM. At Arc concentrations below this value, the single-chain protein is more stable than the unlinked protein, because of an increased refolding rate. The linker has only a small effect on the unfolding rate. The refolding reactions of the covalently linked and unlinked proteins proceed without apparent intermediates and have transition states with similar exposure of nonpolar surface area to solvent. Extrapolation of urea refolding data suggests that the covalently linked protein refolds in a reaction with a half-time of ≈70 μs in the absence of denaturant.

MATERIALS AND METHODS

Spectroscopy. Arc-L1-Arc was expressed in *Escherichia coli* strain X90/λO/pLA110 and purified as described (Robinson & Sauer, 1996). Arc-*st11* was purified from *E. coli* strain X90/λO/pSA700 as described (Milla et al., 1993). Arc-

† Supported by NIH Grant AI-15706 and an NIH/NRSA post-doctoral fellowship (GM-61727) to C.R.R.
⊗ Abstract published in *Advance ACS Abstracts,* October 15, 1996.

S0006-2960(96)01375-X CCC: $12.00

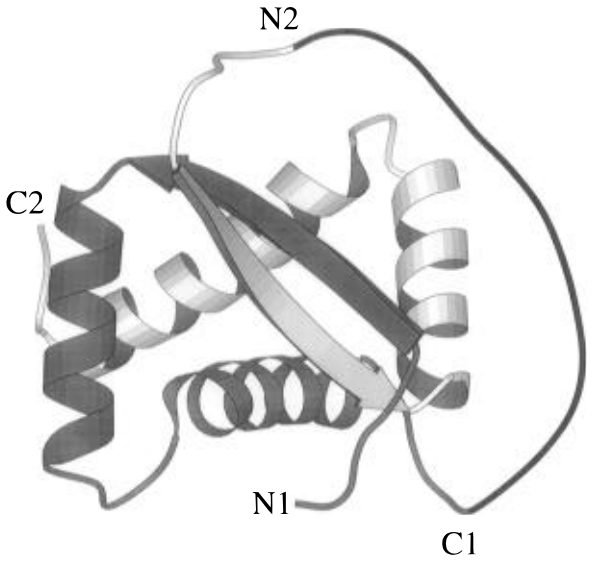

FIGURE 1: Ribbon-trace model of the Arc-L1-Arc protein based on the structure of the DNA-bound form of the Arc dimer (Raumann et al., 1994). The first subunit is colored dark gray, and the second is light gray. The conformation of the L1 linker (black) is one of several possible. The positions of the N- and C-termini of each subunit are indicated. The figure was generated using MOLSCRIPT (Kraulis, 1991).

st11 differs from wild-type Arc in having an 11-residue C-terminal extension (His_6-Lys-Asn-Gln-His-Glu) which prevents intracellular proteolysis and allows affinity purification but does not perturb the stability, activity, or folding kinetics of the protein (Milla et al., 1993). This same 11-residue sequence is also present at the C-terminus of Arc-L1-Arc. Protein concentrations in monomer equivalents were determined using extinction coefficients at 280 nm of 6756 M^{-1} cm^{-1} for Arc-*st11* (Brown et al., 1990) and 13 512 M^{-1} cm^{-1} for Arc-L1-Arc. Thus, 10 μM Arc-*st11* is equivalent to 5 μM Arc-L1-Arc in terms of subunit concentration. UV absorbance was determined using a Hewlett Packard 8452A diode array spectrophotometer. Fluorescence spectra were measured using an Hitachi F-4500 fluorescence spectrophotometer and excitation wavelengths of 280 nm for intrinsic fluorescence of Trp 14, and 360 nm for bis-ANS fluorescence. Circular dichroism (CD) spectra were measured using an AVIV 60DS spectrophotometer.

Equilibrium Measurements. For urea denaturation experiments, two protein samples of equal concentration were prepared in buffer containing 50 mM Tris-HCl (pH 7.5), 250 mM KCl, 0.1 mM EDTA, and either 0 or 10 M urea. One sample was placed in a spectrophotometer or fluorimeter cell equipped with a stirrer and thermostated at 25 °C. By replacing an aliquot from the cell with an equal volume of the other sample, the urea concentration was changed sequentially. At each concentration, the sample was equilibrated until there was no detectable change in signal (typically 1 min), and the CD signal at 222 nm or the fluorescence emission at 337 nm was averaged for 30 s. Denaturation of Arc-L1-Arc was reversible and fit as a two-state transition between the native (N) and denatured (D) states (K_u = [D]/[N]). Denaturation of Arc-*st11* was fit as a two-state transition between the native dimer (N_2) and two denatured (D) monomers (K_u = $[D]^2/[N_2]$) as described (Bowie & Sauer, 1989; Milla et al., 1993). In both cases, the urea dependence of ΔG_u was modeled as a linear function of denaturant:

$$\Delta G_u = \Delta G_u^{H_2O} - m[\text{urea}] \quad (1)$$

where $\Delta G_u^{H_2O}$ is the free energy of denaturation in the absence of denaturant. Equilibrium denaturation data were fit using a Macintosh version of the program NonLin (Johnson & Frasier, 1985; Brenstein, 1989). For Arc-*st11*, the expected relationship between C_m (the urea concentration required for 50% denaturation) and P_t (the total protein concentration) is given by

$$C_m = \frac{2.3RT}{m}\log(P_t) + \frac{\Delta G_u^{H_2O}}{m} \quad (2)$$

where R is the gas constant and T is temperature (K). This expression can be derived from eq 1 after substituting C_m for the urea concentration and $-2.3RT\log(P_t)$ for ΔG_u.

Binding constants for bis-ANS were determined at 25 °C in 50 mM Tris-HCl (pH 7.5), 250 mM KCl, and 0.1 mM EDTA by titrating 0.1 μM solutions of bis-ANS with Arc-*st11* or Arc-L1-Arc protein in a thermostated, stirred sample cell. After each addition, the sample was allowed to equilibrate for 120 s until there was no detectable change in fluorescence, and the fluorescence emission intensity at 490 nm was averaged for 30 s. To determine the apparent K_d, binding isotherms were fit to a hyperbolic function using a subroutine in the program KaleidaGraph.

Kinetic Measurements. Kinetic measurements of the Arc-L1-Arc and Arc-*st11* refolding and unfolding reactions were performed using an Applied Photosystems SpectraKinetic stopped-flow fluorescence instrument. Experiments were performed in a thermostated chamber at 25 °C in buffer containing 50 mM Tris-HCl (pH 7.5), 250 mM KCl, and 0.1 mM EDTA. For Arc-L1-Arc, refolding reactions were initiated by mixing protein denatured in 7.0 M urea with solutions containing 0−2.5 M urea, in 1:5 ratio to give final refolding conditions of 1.2−3.3 M urea. For Arc-*st11*, refolding reactions were initiated by mixing protein denatured in 5.0 M urea with solutions containing 0−2.6 M urea, in 1:5 ratio to give final refolding conditions of 0.8−3.0 M urea. In unfolding experiments, Arc-L1-Arc or Arc-*st11* in buffer without urea was mixed 1:5 with solutions containing 8.0−10.0 M urea to give final concentrations of 6.7−8.3 M urea. In all experiments for Arc-L1-Arc and most experiments for Arc-*st11*, the starting and ending urea concentrations were in the baseline regions of the denaturation reaction, where the protein population is greater than 95% native or denatured. Following mixing, fluorescence was measured by excitation at 280 nm, with emission monitored using a 300 nm cutoff filter. Kinetic data for unfolding of both proteins and refolding of Arc-L1-Arc were fit as single exponentials using the SpectraKinetic software or KaleidaGraph. Data points within the mixing dead time of the instrument (~1 ms) were not used in the fit. Kinetic data for refolding of Arc-*st11* were fit as hyperbolic functions as described (Milla & Sauer, 1994).

RESULTS

Characterization of Native and Denatured States. Three spectroscopic probes were used to compare the native and denatured states of single-chain Arc-L1-Arc with those of Arc-*st11*, an unlinked variant similar to wild-type Arc. As shown in Figure 2, the circular dichroism (CD) and intrinsic

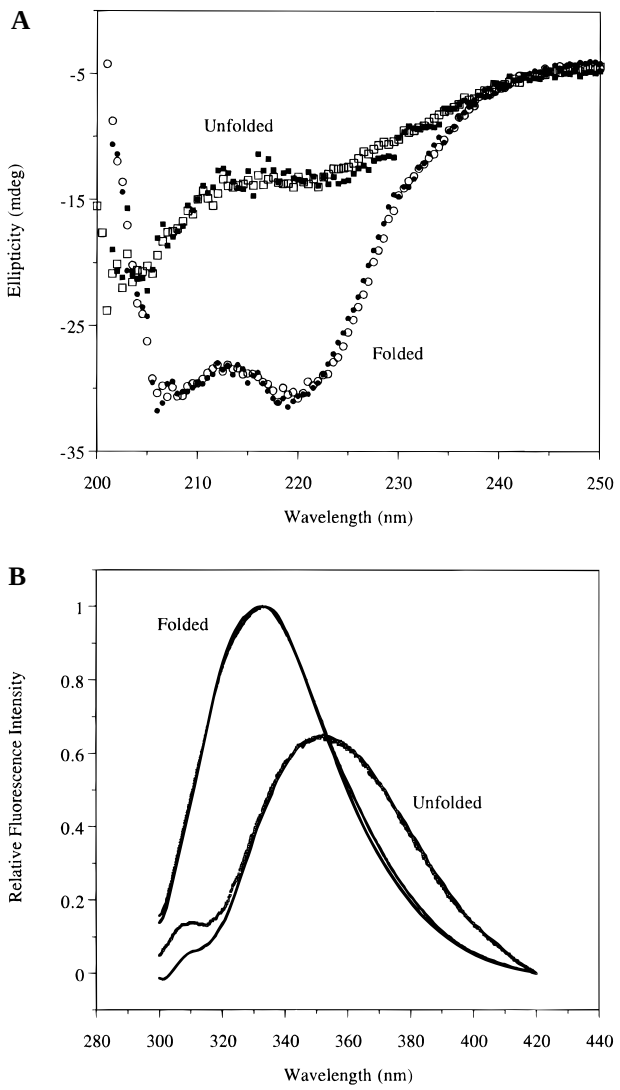

FIGURE 2: (A) Circular dichroism spectra of 5 μM Arc-L1-Arc folded (25 °C, filled circles) and unfolded (90 °C, filled squares) and 10 μM Arc-*st11* folded (25 °C, open circles) and unfolded (90 °C, open squares). (B) Fluorescence spectra of 5 μM Arc-L1-Arc in 0 M urea (folded) and 9 M urea (unfolded). The solid lines are spectra for folded and unfolded Arc-st11 (10 μM) under the same conditions. Measurements were performed in buffer containing 50 mM Tris-HCl (pH 7.5 at 25 °C), 250 mM KCl, and 0.1 mM EDTA. Note that at 90 °C the pH decreases to about 6.0 but that Arc stability does not vary appreciably between pH 6 and 8 (Bowie & Sauer, 1989).

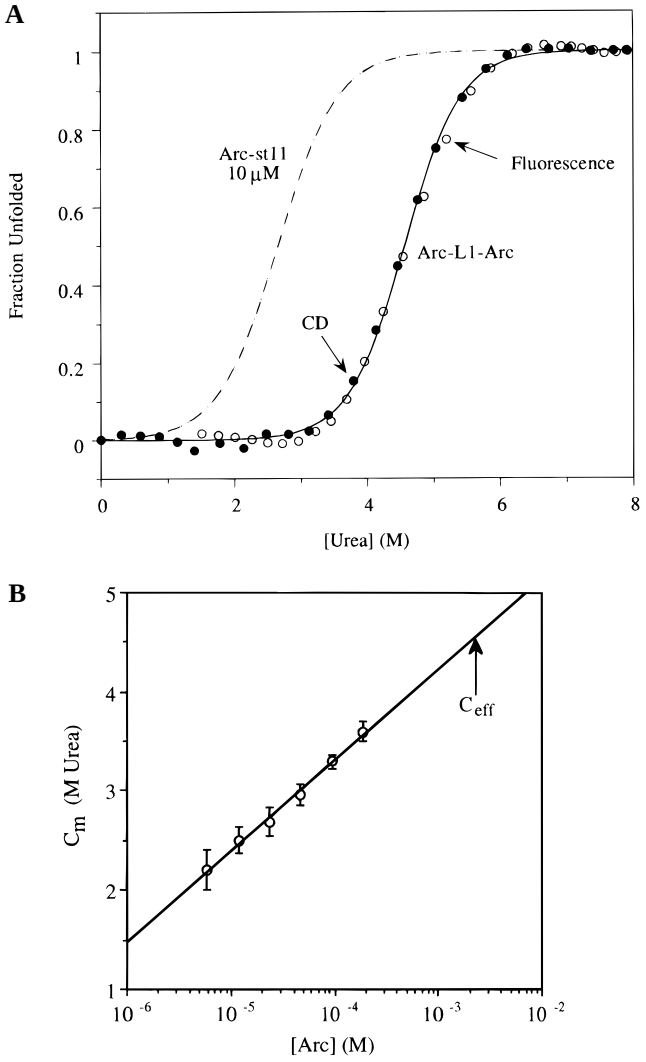

FIGURE 3: (A) Urea unfolding profiles of Arc-L1-Arc (5 μM) monitored by CD ellipticity at 222 nm (closed circles) and fluorescence emission at 337 nm (open circles). The solid line indicates the best fit with $\Delta G_u^{H_2O} = 6.4$ ($\pm$ 0.5) kcal/mol, $m = 1.4$ ($\pm$ 0.1) kcal/mol·M. The dashed line represents the urea unfolding profile for 10 μM Arc-*st11*. Experiments were performed at 25 °C in buffer containing 50 mM Tris-HCl (pH 7.5), 250 mM KCl, and 0.1 mM EDTA. (B) Concentration dependence of C_m for Arc-*st11*. C_m values were determined by urea unfolding experiments monitoring change in fluorescence emission at 337 nm as described (Bowie & Sauer, 1989). The solid line indicates the best fit to the semilogarithmic function $C_m = 2.3RT/m \log(P_t) + \Delta G_u^{H_2O}/m$ with $\Delta G_u^{H_2O} = 10.3$ ($\pm$ 0.5) kcal/mol and $m = 1.48$ ($\pm$ 0.06) kcal/mol·M. Based on these values, the arrow indicates the concentration of Arc-*st11* (2.7 mM) required to give a C_m equivalent to that of Arc-L1-Arc (4.6 M). This represents the effective concentration of subunits (C_{eff}) in the single-chain protein.

fluorescence spectra of native and denatured Arc-L1-Arc are very similar to those of Arc-*st11* at equivalent subunit concentrations. The binding of the native and denatured forms of both proteins to bis-ANS, a hydrophobic dye which binds to exposed hydrophobic patches in proteins, was also assayed. Hydrophobic dyes such as bis-ANS and ANS typically bind to partially folded proteins, but they have been observed to bind with modest affinity to native Arc repressor (Silva et al., 1992; Milla & Sauer, 1995). Bis-ANS binds with similar affinities to the native form of both proteins [$K_d = 7$ ($\pm$ 2) μM for Arc-L1-Arc and 10 ($\pm$ 4) μM for Arc-*st11*] but did not bind detectably to either denatured protein (data not shown). Overall, these data suggest that the native and denatured conformations of Arc-L1-Arc resemble those of the Arc-*st11* protein and thus that the 15 residue linker plays a relatively passive structural role.

Stability at Equilibrium. Changes in circular dichroism and intrinsic fluorescence of Trp 14 were used to monitor the urea denaturation of Arc-L1-Arc. As shown in Figure 3A, both physical probes give essentially the same denaturation curve, as expected for a reaction in which loss of α-helicity and exposure of buried tryptophans to solvent occur concurrently because most protein molecules are either fully native or fully denatured. Fitting the combined fluorescence and CD denaturation data for Arc-L1-Arc gives a value of $\Delta G_u^{H_2O}$ of 6.4 ($\pm$ 0.5) kcal/mol and an *m*-value of 1.40 ($\pm$ 0.10) kcal/mol·M. Fitting the CD and fluorescence melting curves independently gives similar values within error. In the experiment shown in Figure 3A, the

Table 1: Stability Parameters for Equilibrium Urea Denaturation of Arc-L1-Arc[a]

protein concentration (mM)	$\Delta G_u^{H_2O}$ (kcal/mol)	m (kcal/mol·M)	C_m (M urea)
0.5	6.3 ± 0.4	1.36 ± 0.08	4.65 ± 0.09
5.0	6.4 ± 0.5	1.40 ± 0.10	4.60 ± 0.20
10.0	6.2 ± 0.3	1.34 ± 0.05	4.63 ± 0.09

[a] Denaturation experiments like those shown in Figure 3A were performed at 25 °C in 50 mM Tris-HCl (pH 7.5), 250 mM KCl, and 0.1 mM EDTA.

Arc-L1-Arc concentration is 5 μM and the urea concentration (C_m) required for 50% denaturation is 4.6 (± 0.2) M. Urea denaturation experiments performed at Arc-L1-Arc concentrations ranging from 0.5 to 10 μM gave similar values of $\Delta G_u^{H_2O}$, m, and C_m (Table 1). The concentration independence of Arc-L1-Arc denaturation is expected for a reaction in which the native and denatured states are both monomeric.

Denaturation of wild-type Arc, unlike that of Arc-L1-Arc, is concentration dependent (Bowie & Sauer, 1989). As shown in Figure 3B, C_m values determined for Arc-*st11* denaturation at different total protein concentrations vary in a linear fashion ($r^2 = 0.995$) with $\log(P_t)$ as expected from eq 2. A linear regression fit of these data gives a slope ($2.3RT/m$) of 0.92 ± 0.04 M and intercept ($\Delta G_u^{H_2O}/m$) of 7.0 ± 0.2 M, corresponding to an m-value of 1.48 (± 0.06) kcal/mol·M and a $\Delta G_u^{H_2O}$ value of 10.3 (± 0.5) kcal/mol at a standard state concentration of 1 M. These m and $\Delta G_u^{H_2O}$ values are within error of those derived from nonlinear least squares fits of individual urea denaturation experiments and those reported previously for Arc-*st11* under similar temperature and buffer conditions (Milla et al., 1993). The similarity in m-values for denaturation of Arc-L1-Arc and Arc-*st11* also suggests that similar amounts of nonpolar surface become solvent inaccessible during folding of these proteins.

Unfolding Rates. Figure 4A shows the unfolding trajectory for Arc-L1-Arc following a jump from 0 to 7.3 M urea (4.4 μM protein, 25 °C, pH 7.5, 250 mM KCl, PMT voltage = 990). These kinetic data are fit well by a single exponential with an unfolding rate constant (k_u) of 40 s^{-1}. Figure 4B shows the urea dependence of k_u for Arc-L1-Arc and for Arc-*st11*. At each urea concentration, the rate constant for Arc-L1-Arc unfolding is larger by a factor of approximately 2. Values of m_u (calculated as $RT\ \partial \ln(k_u)/\partial[\text{urea}]$) are similar for the two proteins. The fractional burial of surface area in the transition state for unfolding can be estimated as $1 - (m_u/m)$ (Tanford, 1970; Matouschek & Fersht, 1993). For Arc-L1-Arc and Arc-*st11*, these values are 0.81 (± 0.10) and 0.80 (± 0.07), respectively (Table 2). Thus, covalent linkage of the Arc subunits in Arc-L1-Arc has only a small effect on the unfolding rate and has no significant effect on the position of the transition state along the reaction coordinate defined by solvent accessibility.

Refolding Rates. Figure 5A shows the refolding trajectory for Arc-L1-Arc following a jump from 7.0 to 2.4 M urea (3.1 μM protein, 25 °C, pH 7.5, 250 mM KCl, PMT voltage = 800). The data fit well to a single exponential with a refolding rate constant (k_f) of 240 s^{-1}. Greater than 95% of the expected change in amplitude is observed during the data collection phase (i.e., less than 5% of the expected amplitude change occurs in the dead time). In experiments performed at final urea concentrations of 2.0 and 2.8 M, k_f was found

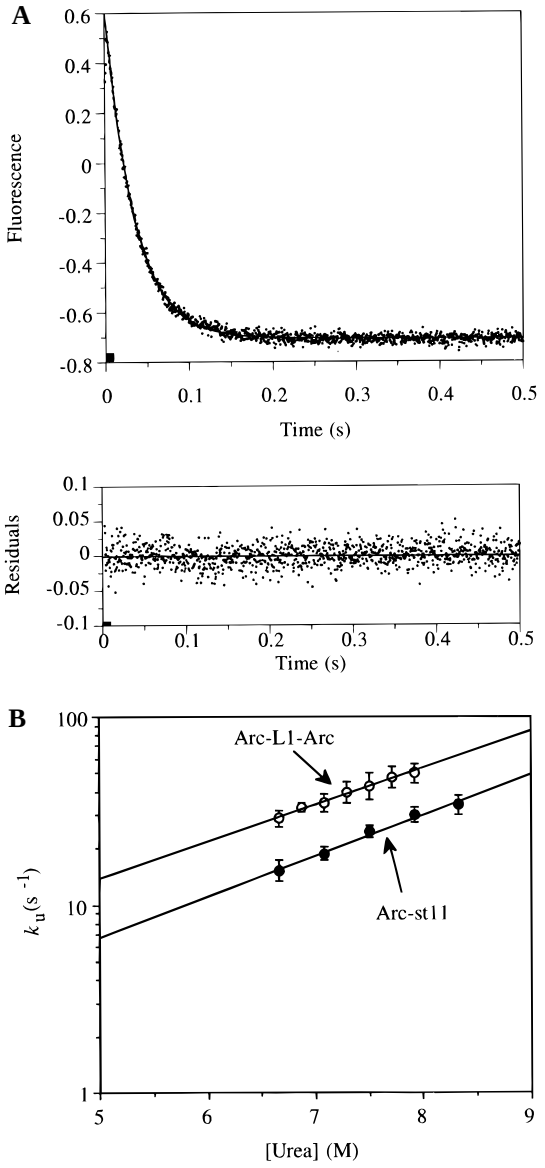

FIGURE 4: Unfolding kinetics. (A) Unfolding trajectory for 4.4 μM Arc-L1-Arc following a jump from 0 to 7.3 M urea. The solid line is the fit to a single exponential with a rate constant of 40 s^{-1}. The lower panel shows the residuals of the fit. Data points within the mixing time (indicated by the solid bar) were not used in the fit. (B) Urea dependence of unfolding rate constants for Arc-L1-Arc (open circles) and Arc-*st11* (filled circles). Error bars indicate one standard deviation from at least five independent experiments. The solid lines are linear regression fits ($r^2 = 0.988$, Arc-L1-Arc; $r^2 = 0.986$, Arc-*st11*). Experiments were performed at 25 °C in buffer containing 50 mM Tris-HCl (pH 7.5), 250 mM KCl, and 0.1 mM EDTA.

to be independent of Arc-L1-Arc protein concentration from 1 to 10 μM (data not shown). Figure 5B shows the refolding rate constants for Arc-L1-Arc plotted as a semilogarithmic function of the urea concentration. For comparison, the apparent rate constants ($k_{app} = k_f P_t$) for refolding of Arc-*st11* at a concentration of 10 μM are also shown. Refolding of Arc-L1-Arc is significantly faster than refolding of Arc-*st11* under these conditions. The urea dependencies of the refolding rate constants are similar for Arc-L1-Arc and Arc-*st11*, and (m_f/m) values, which estimate the fractional burial of surface area in the transition state for refolding, are also within error (Table 2). Linear extrapolation of the data in Figure 5B yields a refolding rate constant of $\sim 10^4$ s^{-1} for Arc-L1-Arc in the absence of urea. This corresponds to a

Table 2: Kinetic and Equilibrium Constants for Refolding and Unfolding of Arc-L1-Arc and Arc-*st11* at 4.6 M Urea[a]

	Arc-L1-Arc	Arc-*st11*	C_{eff}
K_u^{eq}	1.0 ± 0.1	$2.7 \pm 0.6 \times 10^{-3}$ M	2.7 ± 0.7
K_u^{kin}	1.6 ± 0.3	$3.4 \pm 0.8 \times 10^{-3}$ M	2.1 ± 1.0
k_f	7.2 ± 1.0 s^{-1}	$3.2 \pm 0.5 \times 10^{3}$ M^{-1} s^{-1}	4.5 ± 1.8[b]
m_f/m	0.67 ± 0.10	0.69 ± 0.07	
k_u	11.6 ± 1.5 s^{-1}	5.5 ± 1.0 s^{-1}	
$1 - (m_u/m)$	0.81 ± 0.10	0.80 ± 0.07	
ΔG_u^{eq}	0.0 ± 0.1 kcal/mol	3.5 ± 0.4 kcal/mol	
ΔG_u^{kin}	-0.3 ± 0.1 kcal/mol	3.4 ± 0.5 kcal/mol	

[a] Values of the fractional solvent exposure of the transition state for folding (m_f/m) and unfolding ($1 - m_u/m$) were calculated using m_f and m_u values of 0.94 and 0.27 kcal/mol·M, respectively, for Arc-L1-Arc and 1.02 and 0.30 kcal/mol·M for Arc-*st11*. The differences in the fractional exposure of the folding and unfolding transition states (although within experimental error) may also indicate that the position of the transition state changes with urea (Jonsson et al., 1996) since m_u is calculated at high urea concentration and m_f is calculated at low urea concentration. [b] $C_{eff-refold}$ was calculated as described in Results.

refolding reaction with a half-life of 70 μs. As shown in Table 2, the equilibrium denaturation constants (K_u) calculated from kinetic experiments are very similar to those calculated from equilibrium experiments for both Arc-L1-Arc and Arc-*st11*, supporting the idea that denturation behaves as a two-state process for both proteins.

Effective Concentrations. In comparing bimolecular and unimolecular reactions that are otherwise similar, the effective concentration (C_{eff}) is defined as the equilibrium dissociation constant for the bimolecular reaction divided by the equilibrium constant for the unimolecular reaction (Page & Jencks, 1971; Creighton, 1983). For the single-chain Arc system:

$$C_{eff} = \frac{K_u^{Arc-st11}}{K_u^{Arc-L1-Arc}} = \frac{k_u^{Arc-st11}}{k_u^{Arc-L1-Arc}} \frac{2k_f^{Arc-L1-Arc}}{k_f^{Arc-st11}} \quad (3)$$

Table 2 lists kinetic and equilibrium constants at 4.6 M urea and the derived values of C_{eff} under these conditions. C_{eff} calculated from the equilibrium constants is 2.7 ($\pm$ 0.7) mM. C_{eff} calculated from the kinetic constants is 2.1 ($\pm$ 1.0) mM. The effective concentration is also equal to the protein concentration at which C_m for the bimolecular reaction is the same as C_m for the unimolecular reaction. Thus, extrapolation of the data in Figure 3B shows that an Arc-*st11* concentration of 2.7 mM would be required to give a C_m equivalent to that of Arc-L1-Arc.

The effective concentration of Arc-L1-Arc subunits during the refolding reaction ($C_{eff-refold}$) can be calculated by dividing $2k_f$ for Arc-L1-Arc refolding by the second-order rate constant for Arc-*st11* refolding. The factor of 2 is necessary because the Arc-L1-Arc refolding rate is calculated for folding of two subunit equivalents in the single-chain protein. $C_{eff-refold}$, which ranges from roughly 1 mM at 2 M urea to 4.5 mM at 4.6 M urea, is equal to the Arc-*st11* concentration that would be required to achieve a refolding rate equal to $2k_f$ for Arc-L1-Arc. The close similarity in the values of C_{eff} and $C_{eff-refold}$ reinforces the idea that the major effect of the linker on equilibrium stability is manifest in the kinetics of the refolding reaction.

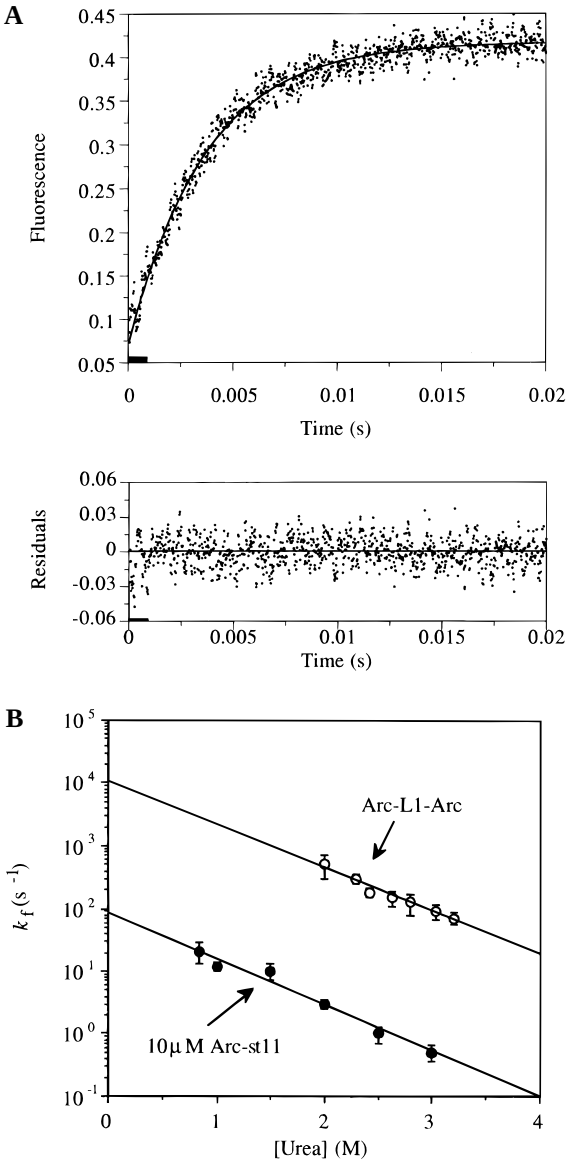

FIGURE 5: Refolding kinetics. (A) Folding trajectory for 3.1 μM Arc-L1-Arc, following a jump from 7.0 to 2.4 M urea. The solid line is the fit to a single exponential with a rate constant of 240 s^{-1}. The lower panel shows the residuals of the fit. Data points within the mixing time (indicated by the solid bar) were not used in the fit. (B) Urea dependence of refolding rate constants for Arc-L1-Arc (open circles) and Arc-*st11* (filled circles). Error bars indicate one standard deviation from at least five independent experiments. The solid lines are linear regression fits ($r^2 = 0.977$, Arc-L1-Arc; $r^2 = 0.975$, Arc-*st11*). Conditions are as in Figure 4.

DISCUSSION

The studies reported here were performed to assess the structural and energetic consequences of connecting the C-terminus of one subunit in the Arc repressor dimer to the N-terminus of the other subunit with the 15-residue linker designated L1. In the design of Arc-L1-Arc, the length of the linker was constrained by the need to span approximately 45 Å, the distance between the termini of two subunits in the Arc dimer when bound to operator DNA (Raumann et al., 1994; Robinson & Sauer, 1996). The L1 sequence (GGGSGGGTGGGSGGG) was chosen in an attempt to maintain maximum flexibility and reasonable solubility, on the basis of studies of single-chain antibodies (Huston et al., 1988; Rumbley et al., 1993).

As assayed by several spectral probes, covalent attachment of the Arc subunits by the L1 linker does not alter the native or denatured protein to any detectable extent. Additionally, the *m*-value for Arc-L1-Arc denaturation is similar to that of dimeric Arc. Because *m*-values provide a measure of the

change in solvent accessible surface between the denatured and native states (Schellman, 1978; Myers et al., 1995), this finding is also consistent with a passive structural role for the linker. Finally, Arc-L1-Arc is fully active in operator binding and repression assays both *in vitro* and *in vivo* (Robinson & Sauer, 1996).

As shown here for the single-chain Arc-L1-Arc protein and previously for the unlinked Arc-*st11* dimer (Milla et al., 1993), denaturation is a reversible reaction which is well described by a two-state model. Since the two equilibrium denaturation reactions differ in molecularity, it is important to specify the protein concentration for the bimolecular reaction when comparing the stabilities of the single-chain and dimeric proteins. At a concentration of 10 μM, for example, more than 50% of the population of Arc-*st11* molecules are denatured at 3 M urea, whereas almost all Arc-L1-Arc molecules are native (Figure 3). The enhanced stability of Arc-L1-Arc, under these conditions, results solely from a refolding rate that is significantly faster than that of Arc-*st11* (Figure 5B); Arc-L1-Arc also unfolds slightly faster than Arc-*st11* (Figure 4B) but this effect, by itself, would decrease the stability of the single-chain protein. The reduced kinetic stability of Arc-L1-Arc makes it unlikely that the linker makes any strong interactions with the rest of the molecule that needed to be broken prior to denaturation. Again, this is consistent with the linker playing a relatively passive structural role.

In the absence of urea, Arc-L1-Arc is predicted to refold with a half-time of 70 μs. To our knowledge, this is the fastest protein folding reaction that has been characterized to date. The N-terminal domain of λ repressor (Huang & Oas, 1995) and the cold-shock protein CspB (Schindler et al., 1995) also have folding reactions that are substantially complete on the sub-millisecond time scale. In each of these cases, the transition from an unfolded polypeptide chain to a native protein with a well-packed hydrophobic core must occur very efficiently. By contrast, other monomeric proteins of similar size often take considerably longer to fold (Matouschek et al., 1990; Jackson & Fersht, 1991; Elöve et al., 1992; Lu & Dahlquist, 1992; Radford et al., 1992). In structural terms, it is not clear why the kinetic folding barriers are low for some proteins and high for others, although recent studies have suggested that formation of buried polar interactions can be one factor that causes proteins of similar size to fold at different rates (Waldburger et al., 1996).

Although the refolding of Arc-L1-Arc is significantly faster than refolding/dimerization of Arc-*st11* at low concentrations, the urea dependence of the rate constants suggests that the transition states for both proteins occur at similar positions along a reaction coordinate defined by solvent accessibility. In terms of burial of hydrophobic surface, the transition states occur relatively late in the refolding reaction (70−80% burial of surface) and relatively early in the unfolding reaction (20−30% exposure of surface). Thus, the molecular events that determine the highest free energy on the reaction coordinate are probably similar for the two folding reactions. Kinetic studies indicate that dimerization of Arc monomers occurs prior to reaching the transition state, which represents a unimolecular barrier between a partially folded dimeric intermediate and the native dimer (Jonsson et al., 1996; Waldburger et al., 1996). Most native interactions appear to be partially formed in this transition state, suggesting that the slow step in folding may involve tight packing of the hydrophobic core, optimization of hydrogen-bond geometries, etc. (Milla & Sauer, 1994; Milla et al., 1995). In this regard, it is not surprising that these same structural events would also be involved in the slow step in Arc-L1-Arc refolding. The linker in Arc-L1-Arc probably acts mainly to hold the denatured chains of each subunit in close proximity during folding, thereby reducing the entropic cost and free energy difference between the denatured state and the transition state. By contrast, during refolding of Arc, the rate at which monomers collide to form the intermediate dimer decreases with decreasing protein concentration. Moreover, because most intermediate dimers dissociate rather than continuing to the native dimer, the overall refolding rate is limited by the entropic cost of this bimolecular pre-equilibration step (Jonsson et al., 1996; Waldburger et al., 1996).

Table 3: Effective Concentrations of Covalently Attached Subunits

protein	C_{eff} (mM)	method	reference
Arc-L1-Arc	2.7	urea unfolding	this work
ROP	0.05[a]	thermal unfolding	Predki and Regan (1995)
Gene V	0.1[b]	urea unfolding	Liang et al. (1993)
Oct1-POU	3.6	DNA binding	Klemm and Pabo (1996)

[a] Extrapolated from concentration dependence of T_m of wild-type ROP. [b] Calculated from comparison of equilibrium urea unfolding data using $P_t = \exp((m_2\Delta G_1/m_1 - \Delta G_2)/RT)$, where m_1 and ΔG_1 are the m and $\Delta G_u^{H_2O}$ values for the subunit-fusion gene V protein, and m_2 and ΔG_2 are these values for wild-type gene V protein.

Effective concentration provides a way of comparing bimolecular and unimolecular reactions that are similar in other respects (Page & Jencks, 1971; Creighton, 1983). By definition,

$$K_d(\text{unimolecular}) = \frac{K_d(\text{bimolecular})}{C_{eff}}$$

Thus, for any given bimolecular reaction, higher values of C_{eff} will result in a unimolecular reaction that proceeds to a greater extent. For Arc-L1-Arc and Arc-*st11*, C_{eff} is 2.7 (± 0.7) mM at 25 °C, pH 7.5, 250 mM KCl, and 4.6 M urea. If, as seems to be the case for Arc-L1-Arc, enthalpic contacts made by the linker are negligible and the structures of the common regions of the linked and unlinked proteins are the same in the denatured and native states, then C_{eff} should represent the protein concentration at which the loss of translational and rotational entropy during refolding and dimerization of Arc-*st11* is equal to the loss of linker entropy during refolding of Arc-L1-Arc. At a concentration of 2.7 mM, the denatured chains of the unlinked Arc-*st11* subunits would be separated by an average distance of roughly 12 Å. In denatured Arc-L1-Arc, the subunits would only need to be this close if the linker were infinitely flexible. Limitations in linker flexibility should increase this value and, in fact, a value of ~20 Å is expected for the end-to-end distance of the L1 linker in a random-coil conformation (Creighton, 1984).

Table 3 lists effective concentrations for other single-chain or covalently attached protein subunits (Liang et al., 1993; Predki & Regan, 1995; Klemm & Pabo, 1996). None of these values is higher than 5 mM, and many are lower. Clearly, these effective concentrations are modest in comparison with effective concentrations (which can be as high

as 10^7 M) in highly constrained small molecules where the unimolecular entropy loss may involve rotation about just one or a few single bonds (Fersht, 1977). This presumably reflects the fact that the intersubunit linkers are quite flexible in both the denatured and native states. It will be interesting, for proteins like Arc, to see if linkers of different length or composition can be designed to give effective concentrations higher than millimolar.

ACKNOWLEDGMENT

We thank Thor Jonsson, Dennis Rentzeperis, and Anne Skaja Robinson for helpful discussions and advice and Jonathan King for use of equipment.

REFERENCES

Bowie, J. U., & Sauer, R. T. (1989) *Biochemistry 28*, 7139−7143.
Brenstein, R. J. (1989) Robelko Software, version 0.9 8b5, Carbondale, IL.
Brown, B. M., Bowie, J. U., & Sauer, R. T. (1990) *Biochemistry 29*, 11189−11195.
Creighton, T. E. (1983) *Biopolymers 22*, 49−58.
Creighton, T. E. (1984) *Proteins*, 1st ed., W. H. Freeman and Company, New York.
Elöve, G. A., Chaffotte, A. F., Roder, H., & Goldberg, M. E. (1992) *Biochemistry 31*, 6876−6883.
Fersht, A. (1977) *Enzyme Structure and Mechanism*, W. H. Freeman and Co., San Francisco.
Huang, G. S., & Oas, T. G. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 6878−6882.
Huston, J. S., Mudgett-Hunter, M., Tai, M.-S., Novotny, J., Margolies, M. N., Ridge, R. J., Bruccoleri, R. E., Haber, E., Crea, R., & Opperman, H. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 5879−5883.
Jackson, S. E., & Fersht, A. R. (1991) *Biochemistry 30*, 10436−10443.
Johnson, M., & Frasier, S. (1985) *Methods Enzymol. 117*, 301−342.
Jonsson, T., Waldburger, C. D., & Sauer, R. T. (1996) *Biochemistry 35*, 4795−4802.
Klemm, J. D., & Pabo, C. O. (1996) *Genes Dev. 10*, 27−36.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Liang, H., & Terwillinger, T. C. (1991) *Biochemistry 30*, 2772−2782.
Liang, H., Sandberg, W. S., & Terwillinger, T. C. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 7010−7014.
Lu, J., & Dahlquist, F. W. (1992) *Biochemistry 31*, 4749−4756.
Matouschek, A., & Fersht, A. R. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 7814−7818.
Matouschek, A., Kellis, J. J., Seraano, L., Bycroft, M., & Fersht, A. R. (1990) *Nature 346*, 440−445.
Milla, M. E., & Sauer, R. T. (1994) *Biochemistry 33*, 1125−1133.
Milla, M. E., & Sauer, R. T. (1995) *Biochemistry 34*, 3344−3351.
Milla, M. E., Brown, B. M., & Sauer, R. T. (1993) *Protein Sci. 2*, 2198−2205.
Milla, M. E., Brown, B. M., Waldburger, C. D., & Sauer, R. T. (1995) *Biochemistry 34*, 13914−13919.
Myers, J. K., Pace, C. N., & Scholtz, J. M. (1995) *Protein Sci. 4*, 2138−2148.
O'Shea, E. K., Rutkowski, R., & Kim, P. S. (1989) *Science 243*, 538−542.
O'Shea, E. K., Klemm, J. D., Kim, P. S., & Alber, T. (1991) *Science 254*, 539−544.
O'Shea, E. K., Rutkowski, R., & Kim, P. S. (1992) *Cell 68*, 699−708.
Page, M. I., & Jencks, W. P. (1971) *Proc. Natl. Acad. Sci. U.S.A. 68*, 1678−1683.
Predki, P. F., & Regan, L. (1995) *Biochemistry 34*, 9834−9839.
Radford, S. E., Dobson, C. M., & Evans, P. E. (1992) *Nature 358*, 302−307.
Raumann, B. E., Rould, M. A., Pabo, C. O., & Sauer, R. T. (1994) *Nature 367*, 754−757.
Robinson, C. R., & Sauer, R. T. (1996) *Biochemistry 35*, 109−116.
Rumbley, C. A., Denzin, L. K., Yantz, L., Tetin, S. Y., & Voss, E. W., Jr. (1993) *J. Biol. Chem. 268*, 13667−13674.
Schellman, J. A. (1978) *Bioploymers 17*, 1305−1322.
Schindler, T., Herrler, M., Marahiel, M. A., & Schmid, F. X. (1995) *Nat. Struct. Biol. 2*, 663−673.
Silva, J. L., Silveira, C. F., Correia, A., Jr., & Pontes, L. (1992) *J. Mol. Biol. 223*, 545−555.
Skinner, M. M., Zhang, H., Leschnitzer, D. H., Guan, Y., Bellamy, H., Sweet, R. M., Gray, C. W., Konnings, R. N. H., Wang, A. H.-J., & Terwilliger, T. C. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 2071−2075.
Steif, C., Weber, P., Hinz, H.-J., Flossdorf, J., Cesareni, G., & Kokkinidis, M. (1993) *Biochemistry 32*, 3867−3876.
Tanford, C. (1970) *Adv. Protein Chem. 24*, 1−95.
Waldburger, C. D., Jonsson, T., & Sauer, R. T. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 2629−2634.

BI961375T