

This paper was presented at the colloquium “Computational Biomolecular Science,” organized by Russell Doolittle, J. Andrew McCammon, and Peter G. Wolynes, held September 11–13, 1997, sponsored by the National Academy of Sciences at the Arnold and Mabel Beckman Center in Irvine, CA.

Optimizing the stability of single-chain proteins by linker length and composition mutagenesis

CLIFFORD R. ROBINSON* AND ROBERT T. SAUER†

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

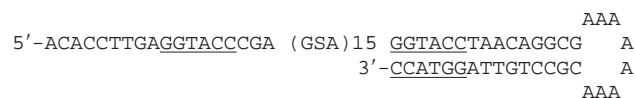
ABSTRACT Linker length and composition were varied in libraries of single-chain Arc repressor, resulting in proteins with effective concentrations ranging over six orders of magnitude (10 μ M–10 M). Linkers of 11 residues or more were required for biological activity. Equilibrium stability varied substantially with linker length, reaching a maximum for glycine-rich linkers containing 19 residues. The effects of linker length on equilibrium stability arise from significant and sometimes opposing changes in folding and unfolding kinetics. By fixing the linker length at 19 residues and varying the ratio of Ala/Gly or Ser/Gly in a 16-residue-randomized region, the effects of linker flexibility were examined. In these libraries, composition rather than sequence appears to determine stability. Maximum stability in the Ala/Gly library was observed for a protein containing 11 alanines and five glycines in the randomized region of the linker. In the Ser/Gly library, the most stable protein had seven serines and nine glycines in this region. Analysis of folding and unfolding rates suggests that alanine acts largely by accelerating folding, whereas serine acts predominantly to slow unfolding. These results demonstrate an important role for linker design in determining the stability and folding kinetics of single-chain proteins and suggest strategies for optimizing these parameters.

The construction of single-chain or hybrid proteins is a potentially powerful method for generating proteins with novel functions and improved properties (1–11). A critical element in such efforts is the design of the peptide linkers that serve to connect different protein domains or subunits. Designed linkers are usually glycine-based peptides with lengths calculated to span the minimum distance between the C terminus of one subunit or domain and the N terminus of the next. How important is linker design in determining the properties of single-chain proteins? Alterations in linker regions have been found to affect the stability, oligomeric state, proteolytic resistance, and solubility of single-chain proteins (12–23), but few systematic investigations of these relationships have been reported. Here, we test the effects of linker design on the stability, protein folding kinetics, and biological activity of single-chain Arc repressor. Wild-type Arc is a dimer with identical subunits, and Arc-L1-Arc is a single-chain variant with a 15-residue linker connecting the subunits (see Fig. 1). The L1 linker of Arc-L1-Arc holds the subunits at an effective concentration (C_{eff}) of 3 mM. By varying linker length and composition, we have isolated single-chain variants with effective subunit concentrations ranging from 10 μ M to 10 M, corresponding to changes in the free energy of unfolding (ΔG_{u}) from 3 to 11 kcal/mol. These differences in stability arise from changes in the folding and unfolding rates, suggest-

ing that linker design can affect protein stability by altering the free energies of both the native and denatured states.

MATERIALS AND METHODS

Cassettes coding for glycine-rich linkers ranging from 3 to 59 residues (Fig. 3A) were synthesized using an Applied Biosystems 381A DNA synthesizer and were purified as described (9). A precursor plasmid (pLA3), constructed to facilitate subcloning of linker library cassettes, contains tandem *arc* genes connected by a GGT ACC GGT adapter, which encodes Gly-Thr-Gly and contains unique *KpnI* and *AgeI* restriction sites. Cassette libraries coding for 19-residue linkers with different amounts of Gly or Ala were constructed by synthesizing an oligonucleotide, which formed a hairpin:



The underlined sequences are *KpnI* sites. S represents a mixture of G and C, and thus, the GSA codons encode either glycine (GGA) or alanine (GCA). Three otherwise identical oligonucleotides with different G/C ratios at the randomized positions (1:1; 3:1; 1:3) were synthesized to facilitate identification of a wide range of compositions. A cassette library encoding random combinations of glycine (GGT) and serine (AGT) was constructed in the same manner. Second strand synthesis was carried out using Sequenase v.2.0 (United States Biochemical) for 2 h at 37°C in Sequenase buffer containing 1 mM dNTPs. Cassettes were digested with *KpnI* and ligated to the *KpnI* backbone of pLA3. Following transformation into *Escherichia coli* strain HB101, colonies were picked randomly and the appropriate region of the single-chain *arc* gene was sequenced using the dideoxy method. Plasmid DNA encoding in-frame constructs were transformed into *E. coli* strain UA2F for assays of activity *in vivo* (24) and into *E. coli* X90- λ O cells for protein expression.

All single-chain Arc proteins contained a (His)₆ tail to facilitate purification using Ni-nitrilotriacetic acid chromatography. Protein purification, fluorescence and circular dichroism (CD) spectroscopy, analytical ultracentrifugation, and gel mobility-shift assays were performed as described (9, 25). Protein stability was assayed by urea denaturation by following changes in intrinsic tryptophan fluorescence intensity at 337 nm or CD ellipticity at 234 nm. For these experiments, the protein concentration was 10 μ M in buffer containing 50 mM

Abbreviations: C_{eff} , effective concentration; CD, circular dichroism.

*Present address: 3-Dimensional Pharmaceuticals, Exton, PA.

†To whom reprint requests should be addressed. e-mail: bobsauer@mit.edu.

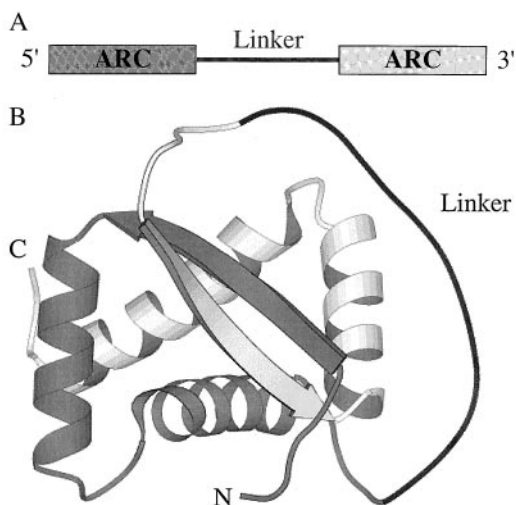


FIG. 1. (A) Tandem copies of the *arc* gene connected by DNA encoding a linker region comprise the gene for single-chain Arc repressor. (B) One model of how a linker might connect the two subunits (colored gray and white) of single-chain Arc. The positions of the N and C termini are indicated. Prepared using MOLSCRIPT (34) and coordinates of wild-type Arc (33).

Tris-HCl (pH 7.5 at 25°C), 250 mM KCl, and 0.1 mM EDTA (26). Values of ΔG_u and m were obtained by fitting denaturation data to a two-state model by nonlinear least squares methods (26). Effective concentrations were calculated by using the equation $C_{\text{eff}} = \exp[(m_2 \bullet \Delta G_1 / m_1 - \Delta G_2) / RT]$, where m_1 and ΔG_1 are values for the single-chain protein, and m_2 and ΔG_2 are values for wild-type Arc (1.48 kcal/mol•M and 10.3 kcal/mol, respectively) (26). Stopped-flow kinetic experiments of protein folding and unfolding were monitored by changes in fluorescence at protein concentrations between 1 and 10 μM in the buffer used for stability measurements (26). Unfolding was initiated by urea-jump experiments (mixing ratio 1:10) to yield a final urea concentration of 7 or 9.1 M. Refolding was initiated by mixing protein denatured in 6.0–9.6 M urea with low urea buffer (1:5 ratio) to yield final urea concentrations between 1.0 and 4.5 M. Rate constants were obtained by fitting the kinetic data to single exponentials. In all cases, the residuals of the fits were distributed randomly. For ease of comparison among each library of variants, rates were either measured at a single urea concentration or measured at a series of urea concentrations and extrapolated to this reference concentration by using linear regression of $\ln(k)$ vs. [urea] plots ($R > 0.99$).

RESULTS

Variation of Linker Length. A library of single-chain *arc* genes with linkers composed of Gly, Ser, and Thr and lengths varying from 3 to 59 aa was constructed (Fig. 3A). The fraction of Gly in different linkers ranges from 66 to 80%. The linkers and corresponding proteins are named LLX and Arc-LLX-Arc (Length Library, X = number of residues), respectively. No intracellular expression of the Arc-LL8-Arc protein was detected. Arc-LL3-Arc expressed to high levels but monomers, dimers, and higher-order oligomers were observed following SDS electrophoresis and Western analysis. This behavior may indicate “cross-folding” as has been observed with single-chain antibodies that have very short linkers (27, 28). The remaining 13 proteins in this library were all expressed at high levels and electrophoresed as monomers. The Arc-LLX-Arc variants were tested for repression of transcription of the P_{ant} promoter in *E. coli* strain UA2F, using resistance to streptomycin as an assay of biological activity (24). Arc-LLX-Arc proteins with

linkers containing 13 or more residues had wild-type activities. Arc-LL11-Arc was partially active; single-chain molecules with the LL3, LL8, and LL9 linkers were inactive. Modeling studies show that connecting the Arc subunits with linkers shorter than 13 residues would either require the linker to cross the DNA-binding surface of the protein and/or require distortion of the structure.

Single-chain Arcs with linkers LL9–LL59 were purified for biophysical characterization. All of these single-chain proteins had CD and fluorescence spectra similar to wild-type Arc. Arc-LL11-Arc, Arc-LL19-Arc, and Arc-LL31-Arc were analyzed by analytical ultracentrifugation and found to be monomeric at concentrations between 10 and 100 μM (data not shown). Proteins containing the three longest linkers (LL47, LL51, and LL59) tended to precipitate at concentrations >100 μM , possibly because of aggregation caused by cross-folding of the Arc subunits.

The thermodynamic stabilities of Arc-LLX-Arc proteins with linkers from 9 to 57 residues were determined by urea denaturation studies, revealing that the 19-residue linker provides maximal stability. As shown in Fig. 2 for a subset of these proteins, there are large changes in the concentration of urea required for denaturation of proteins with different linker lengths, but the curves are roughly parallel indicating that the denaturant m -values (variation of ΔG_u with urea) are similar. Fig. 3B shows the variations of ΔG_u and C_{eff} with linker length. For linkers from 9 to 19 residues, stability of the single-chain protein increased with length. Arc-L9-Arc was the least stable ($\Delta G_u \approx 3$ kcal/mol; $C_{\text{eff}} \approx 6$ μM) and Arc-LL19-Arc was the most stable ($\Delta G_u = 8.4$ kcal/mol; $C_{\text{eff}} = 80$ mM) of the proteins examined. Increases in linker length past 19 residues resulted in decreasing stability until a plateau was reached at ≈ 4.5 kcal/mol ($C_{\text{eff}} \approx 150$ μM) for linkers between 47 and 59 residues.

The linker-dependent changes in stability arise from changes in both the folding and unfolding rates, as measured in urea-jump, stopped-flow, kinetic experiments. Fig. 3C and D show that both the folding and unfolding rate constants vary significantly as the linker length is changed. In 7 M urea, Arc-LL9-Arc unfolds with a rate constant (k_u) of $\approx 3,000$ s^{-1} . As the linker length is increased from 9 to 19, there is a roughly exponential decrease in k_u that spans 3–4 orders of magnitude and reaches a value of ≈ 1 s^{-1} for Arc-LL19-Arc. Changes in linker length between 19 and 59 residues do not change k_u appreciably. Thus, linkers shorter than 19 residues reduce the

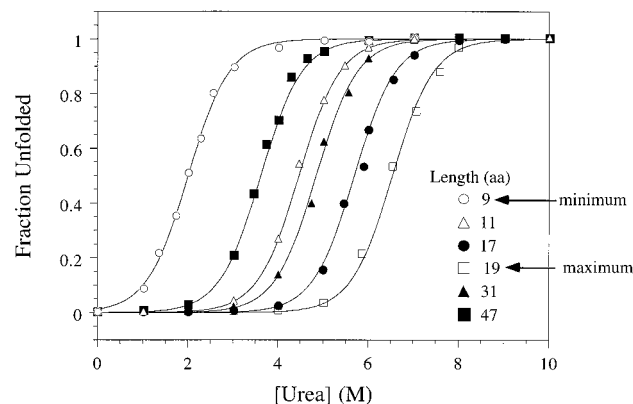


FIG. 2. Linker length has large effects on the stability of single-chain Arc to urea denaturation. The sequences of linkers LL9 (○), LL11 (△), LL17 (●), LL19 (□), LL31 (▲), and LL47 (■) are listed in Fig. 3A. Fraction unfolded was calculated by fitting plots of CD ellipticity (234 nm) vs. urea concentration to a two-state-unfolding transition. The solid lines represent the best theoretical fits of the experimental data.

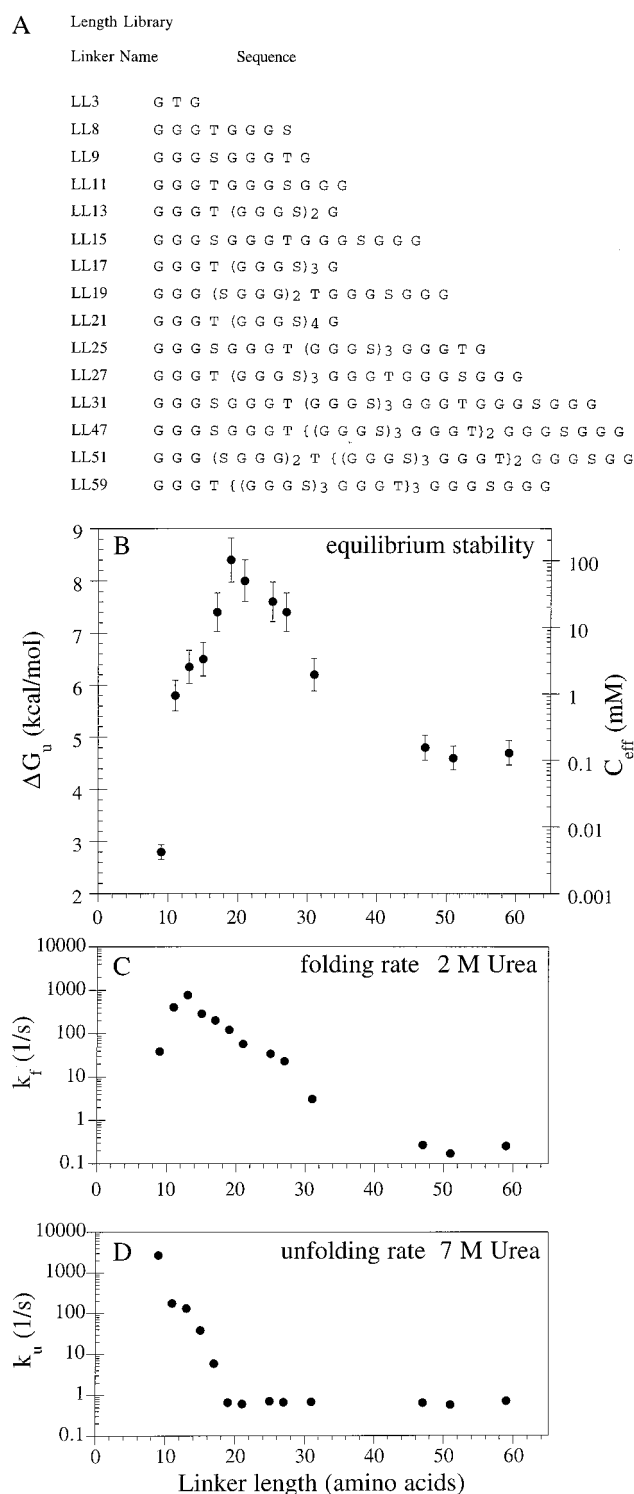


FIG. 3. Properties of linker-length variants of single-chain Arc. (A) Linker sequences. (B) Equilibrium stability and effective concentration vary with linker length. Error bars indicate one SD from three independent experiments. (C) Folding rates in 2 M urea. (D) Unfolding rates in 7 M urea. Experimental conditions; protein 1–10 μ M, 25°C, 50 mM Tris-HCl (pH 7.5), 250 mM KCl, and 0.1 mM EDTA.

free energy barrier between the native state and the transition state.

The refolding rate (k_f) in 2 M urea has a maximum value of $\approx 1000 \text{ s}^{-1}$ for the Arc-LL13-Arc protein. Decreasing the linker by four residues to a length of nine causes a 30-fold decrease in the folding rate. As the linker length is increased from 13 to

47 residues, the refolding rate also decreases. Over this range, there is a roughly exponential decrease in k_f that spans nearly four orders of magnitude. Little change in k_f is seen for linkers between 47 and 59 residues. These results show that linker length can have large effects on the free energy difference between the denatured state and the transition state. Moreover, the length optima for equilibrium stability (19 residues), refolding (13 residues), and unfolding (19–59 residues) are different. The 19-residue linker provides the greatest equilibrium stability because it is the best compromise between reasonably fast refolding and slow unfolding.

Effects of Linker Composition. To assess the effects of varying the number of glycines in the linker, the length of the linker was fixed at 19 residues and 16 internal positions were randomized between Ala and Gly (ALX library) or between Ser and Gly (SLX library) by using the strategy described in *Materials and Methods*. For these experiments, the libraries were first selected for Arc repressor activity *in vivo* and then the sequences of individual members were determined. Sixteen proteins comprise the ALX library; the linkers in these proteins contain from 3 to 15 alanines (Fig. 4A). Ten proteins, with 3–11 serines in the linker region, comprise the SLX library (Fig. 5A). All of the Arc-ALX-Arc and Arc-SLX-Arc proteins were expressed at high levels, were purified, and had CD and fluorescence spectra similar to wild-type Arc. In the ALX library, variants with eight or more linker alanines showed some tendency to aggregate during purification and handling but were monomeric at concentrations of 1–20 μ M as judged by analytical ultracentrifugation and the concentration independence of equilibrium stability and refolding rates. All other proteins in the ALX and SLX libraries were highly soluble.

The number of non-glycine residues in the 19-residue linker has a significant effect on the equilibrium stability of proteins in both the ALX and SLX libraries, as determined by urea denaturation. In the ALX library (Fig. 4A and B), Arc-AL11-Arc, which contains 11 alanines and 5 glycines in the randomized portion of the linker, has the maximum stability ($\Delta G_u \approx 11 \text{ kcal/mol}$; $C_{eff} \approx 8 \text{ M}$). Arc-AL3-Arc, with 3 alanines and 13 glycines in the randomized region of the linker, is far less stable ($\Delta G_u \approx 3 \text{ kcal/mol}$; $C_{eff} \approx 10 \mu\text{M}$), suggesting that too much linker flexibility is detrimental to stability. Fig. 4B shows, however, that stability also decreases when the number of alanines is increased past the optimum value of 11, indicating that linkers that are too inflexible also limit protein stability. The same general trends are observed in the SLX library; proteins with too many or too few glycines are significantly less stable than Arc-SL7-Arc ($\Delta G_u \approx 7 \text{ kcal/mol}$; $C_{eff} \approx 7 \text{ mM}$). There are, however, two significant differences between the ALX and SLX results. Maximum stability occurs for a protein containing eight glycines in the randomized portion of the linker in the SLX library but for a protein containing only five glycines in this region in the ALX library. Moreover, the stabilities of the most stable variants in each library also differ significantly; Arc-AL11-Arc has an effective concentration that is 1,000-fold greater than Arc-SL7-Arc. We interpret these differences as indicating that the identity of the non-glycine residues in the linker is as important as the number of these residues in determining stability. By contrast, the positions of the glycine and non-glycine residues in the randomized portion of the linker seem to be unimportant. Five pairs of variants in the ALX library and three pairs in the SLX library have the same composition but difference sequences. In each of these cases, the stabilities of these variants (indicated by open and closed symbols in Figs. 4B and 5B) were found to be within experimental error.

Another significant difference between the ALX and SLX libraries is observed in the unfolding kinetics (Figs. 4D and 5D). In the ALX library, the unfolding rate of different variants only changes by a factor of 20. In the SLX library, the unfolding rates change by $>1,000$ -fold. In addition, the shapes

A Alanine Library

Linker Name	Sequence
AL3	G T A G G G G G G A G G A G G G G G G
AL4	G T A G G G G G G G A G G A G A G G
AL6a	G T A G A G G A G G G A G G A G A G
AL6b	G T A G G A G A G G G A G A A G G G
AL7	G T A A G G A G A G G G G A G G A G G
AL8a	G T A G G A A G A G G A G A G A A G G
AL8b	G T A A G A G G A G G G A G G A A A G
AL9a	G T A G A G G G A G G A A A A G A A G
AL9b	G T A G A A G A G A A G A A A G G A G
AL10	G T A A G A G G A A A G A G A A G A G
AL11	G T A A A G A G A A G G A A A G A A G
AL12	G T A A A A G A A A G A A G A A G A G
AL14a	G T A A A A G A G A A A A A A A A A G
AL14b	G T A A A A A A A A G A A A G A A A G
AL15a	G T A A A A A G A A A A A A A A A A G
AL15b	G T A A A A A A A A A A A A A A G A A G

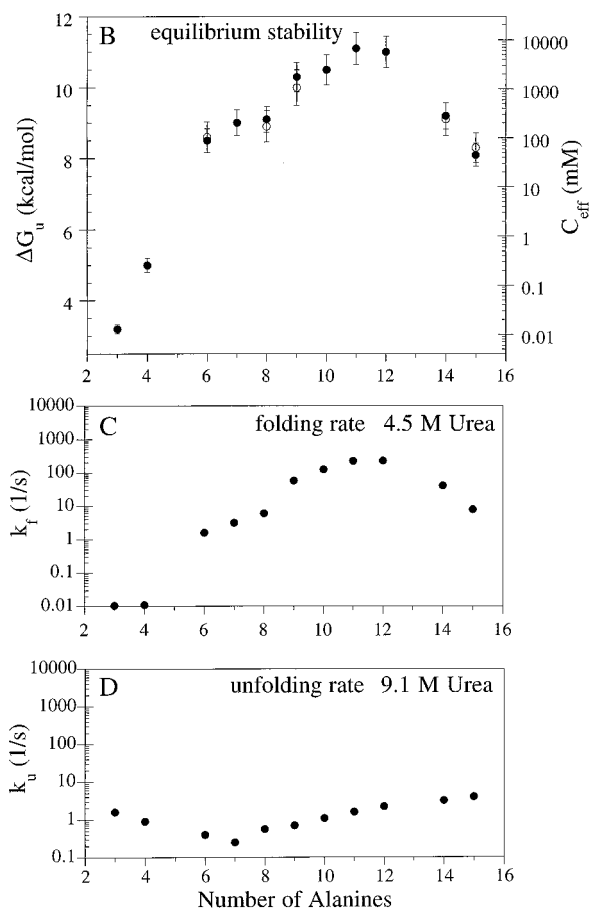


FIG. 4. Properties of ALX variants with 19-residue linkers and differing in Ala/Gly composition numbers of alanines and glycines. (A) Linker sequences. (B) Equilibrium stability and effective concentration vary with number of alanines. For compositional isomers, closed and open symbols represents "a" and "b" variants, respectively. Error bars indicate one SD from three independent experiments. (C) Folding rates in 4.5 M urea. (D) Unfolding rates in 9.1 M urea. See Fig. 3 for conditions.

of these plots are very different. The ALX data is concave upward with minimum occurring for the protein with seven

A Serine Library

Linker Name	Sequence
SL3	G T S G G G S G G G G G G S G G G G
SL4	G T S G G G S G G G S G G S G G G G
SL5a	G T S S G G S G G G G S G S G G G G
SL5b	G T S G G G S G G S G S G S G G G G
SL7	G T S G S S G S G S G G S G S G G G G
SL8a	G T S S G S G S S G G G S G S G G S G
SL8b	G T S S G G S G S S G S G S G G S G G
SL10a	G T S S G S S G S S G G S S S G S G G
SL10b	G T S S S G S G G S G S S G G S S S G
SL11	G T S G S G S S S G S S S G G S S S S G

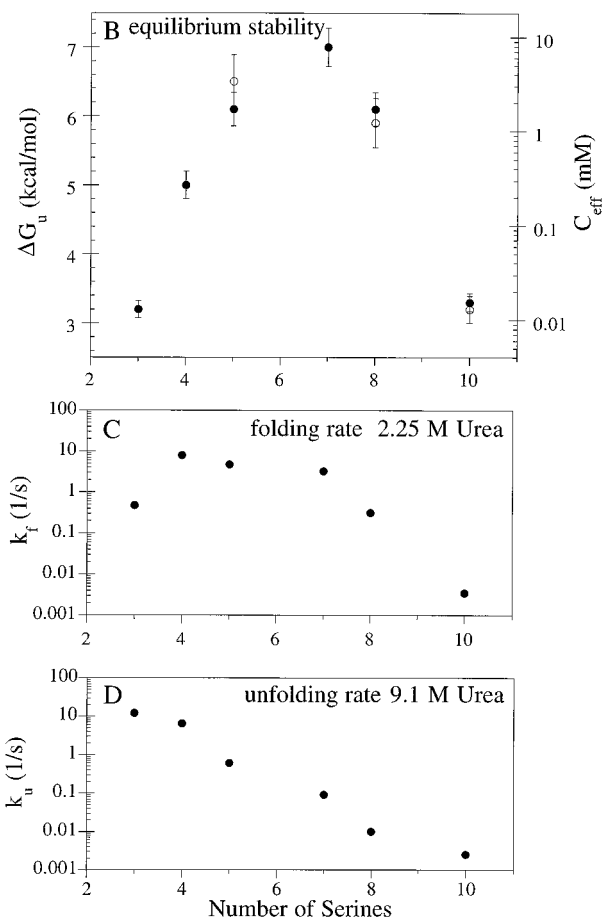


FIG. 5. Properties of SLX variants with 19-residue linkers differing in Ser/Gly composition. (A) Linker sequences. (B) Equilibrium stability and effective concentration vary with number of serines. For compositional isomers, closed and open symbols represents "a" and "b" variants, respectively. Error bars indicate one SD from three independent experiments. (C) Folding rates in 2.25 M urea. (D) Unfolding rates in 9.1 M urea. See Fig. 3 for conditions.

alanines and eight glycines in the randomized portion of the linker. In the SLX library, by contrast, k_u decrease exponentially with the number of serines. The rate constants for refolding in the ALX library change by more than five orders of magnitude, reaching a maximum for variants with 11 or 12 alanines in the randomized part of the linker (Fig. 5C). Because changes in the unfolding rate are small for the ALX proteins, the changes in equilibrium stability arise almost exclusively from changes in the refolding rate. In the SLX library, variants differ over a 300-fold range in refolding rates with a maximum between four and seven serines. Because

much larger changes are seen in the unfolding rates, the changes in equilibrium stability for the SLX proteins are dominated by the changes in unfolding kinetics. These results emphasize once again that the chemical identity of the non-glycine residues in the linker can have a profound effect on the biophysical properties of the single-chain proteins.

DISCUSSION

Linker length and composition exert a surprisingly large influence on the stability of single-chain Arc repressor. In the LLX linker length library, the most stable protein has a linker of 19 residues, and adding or deleting a few amino acids decreases stability (Fig. 3B). These length effects on stability arise from changes in the folding and unfolding rates. In the regime from 59 to 13 residues, shortening the linker accelerates folding. This observation is explained most simply if the denatured subunit domains are constrained to smaller and smaller regions of conformation space by shorter linkers and thus require less random sampling before essential collisions required for folding occur. We note, however, that the length dependence of the stability of single-chain Arc variants in this regime is significantly steeper than for loop-length variants of single-chain Rop (29) and is modeled poorly by simple, random walk, entropic considerations (30). As the linker length decreases from 13 to 11 to 9 residues, there is a decrease in the folding rate of the corresponding Arc-LLX-Arc protein. At some point, the linkers must become too short to connect the subunits in the native conformation without strain. In fact, in the linker length regime from 19 to 9 residues, the unfolding rates of the corresponding Arc-LLX-Arc proteins increases exponentially as the linkers become shorter, suggesting that shorter tethers in this length range introduce more and more strain into the native structure. Presumably, proteins with the LL17, LL15, and LL13 linkers do not show decreased folding rates because of compensating changes in conformational search efficiency.

Glycine is generally used in designed linkers because the absence of a β -carbon permits the polypeptide backbone to access dihedral angles that are energetically forbidden for other amino acids (31). Thus, a glycine-rich linker will be more flexible than a linker of comparable length composed of non-glycine residues. Our results, however, indicate that too much linker flexibility is detrimental to single-chain protein stability. In the ALX (alanine/glycine) library, maximum stability was observed when the 16-residue-randomized region contained 11 alanines and 5 glycines. In the SLX (serine/glycine) library, the most stable protein had seven serines and nine glycines in the randomized portion of the linker. In both libraries, plots of stability vs. the number of non-glycine residues are relatively regular and proteins with the same linker compositions have comparable stabilities (Figs. 4B and 5B). Both observations suggest that it is the composition rather than the sequence of the linker that is important in determining stability. A single exception to this generalization is provided by Arc-LL19-Arc and Arc-SL3-Arc, which have the same composition but stabilities differing by 3.4 kcal/mol. The first three residues of the linker are Gly-Thr-Ser in Arc-SL3-Arc, which has lower stability, and Gly-Gly-Gly in Arc-LL19-Arc, suggesting that the conformational flexibility imparted by glycine may be important at the junction between C terminus of the first subunit and the N terminus of the linker.

In the ALX library, the main effects of alanine composition on stability result from changes in the refolding rate. For example, as the number of alanines in the linker increases from 3 to 11, the folding rates of the corresponding proteins increase by 30,000-fold. Alanine restricts the number of allowed conformations of the linker compared with glycine and, in this length regime, probably accelerates the conformational search that occurs during folding. Increasing the number of alanines

to 14 or 15 then reduces the folding rate, probably because these linkers become too inflexible. When serine is substituted for glycine, there are also effects on the refolding rate but with several differences: the optimal number of serines is smaller than the optimal number of alanines (7 Ser vs. 11 Ala), the difference between the fastest and slowest folders are smaller ($\approx 2,000$ -fold for SLX vs. $\approx 30,000$ -fold for ALX), and the maximum folding rates are different (in 2.25 M urea, the fastest ALX protein folds ≈ 250 times faster than the fastest SLX protein). Clearly, alanine and serine affect linker flexibility in rather different ways.

Large differences between alanine and serine are also apparent when comparing effects on the unfolding rate. As the number of serines in the linker increases, the unfolding rate continues to decrease over a 5,000-fold range (Fig. 5D). By contrast, in the alanine library, the minimum unfolding rate is observed for a protein with seven alanines and the total change between the slowest and fastest unfolders is only 15-fold. We presume that the ability of serine to form hydrogen bonds allows formation of new stabilizing interactions in the native state but whether these interactions are within the linker or involve interactions between the linker and the body of the single-chain protein is unknown. Because alanines in the linker primarily affect folding rates whereas serine has the largest effects on unfolding rates, it seems possible that optimizing the composition of Gly, Ser, and Ala in a linker library might produce single-chain molecules with even greater stabilities than those described here. Preliminary studies also suggest that the effects of length and composition may be interdependent. For example, linkers of different lengths may have different optimal compositions.

Variations in linker length or composition caused no significant changes in repressor activity *in vivo* except in proteins with linkers shorter than 11 residues. In gel mobility-shift assays, Arc-LL19-Arc and Arc-LA11-Arc, which have 19-residue linkers, bound operator DNA as strongly as wild-type Arc dimers (data not shown). In earlier work, however, we found that Arc-L1-Arc (which is identical to Arc-LL15-Arc) had a 10-fold enhanced affinity for operator DNA (9, 26). In single-chain Arc, the linker connects the N-terminal arm of the second subunit to the C terminus of the first subunit; in wild-type Arc, this N-terminal arm is disordered in solution (32) but folds against the operator in the protein-DNA complex (33). The L1/LL15 linker may increase operator affinity by helping to restrict the conformation of the arm in solution, thereby reducing the entropic penalty for ordering the arm upon DNA binding (9). By this model, lengthening the linker to 19 residues probably reduces constraints on the arm conformation.

In summary, we find that changes in linker length and composition can produce substantial changes in the stability and folding kinetics of single-chain Arc. Poly-glycine linkers maximize the conformational freedom of the polypeptide backbone but do not result in optimal stability. For single-chain or hybrid protein designs that have folding problems, alterations in linker length and/or composition should provide a useful method for increasing stability.

We thank David Goldenberg for helpful discussions. This work was supported by an National Institutes of Health postdoctoral fellowship (to C.R.R.) and by National Institutes of Health Grant AI-15706 (to R.T.S.).

1. Bird, R. E., Hardman, K. D., Jacobson, J. W., Johnson, S., Kaufman, B. M., Lee, S.-M., Lee, T., Pope, S. H., Riordan, G. S. & Whitlow, M. (1988) *Science* **242**, 423–426.
2. Pomerantz, J. L., Sharp, P. A. & Pabo, C. O. (1995) *Science* **267**, 93–96.
3. Predki, P. F. & Regan, L. (1995) *Biochem.* **34**, 9834–9839.

4. Hallewell, R. A., Laria, I., Tabrizi, A., Carlin, C., Getzoff, E. D., Tainer, J. A., Cousens, L. S. & Mullenbach, G. T. (1989) *J. Biol. Chem.* **264**, 5260–5268.
5. Bizub, D., Weber, I. T., Cameron, C. E., Leis, J. P. & Skalka, A. M. (1991) *J. Biol. Chem.* **266**, 4951–4958.
6. Kim, S.-H., Kang, C.-H., Kim, R., Cho, J. M., Lee, Y.-B. & Lee, T.-K. (1989) *Protein Eng.* **2**, 571–575.
7. Liang, H., Sandberg, W. S. & Terwillinger, T. C. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7010–7014.
8. Toth, M. J. & Schimmel, P. (1986) *J. Biol. Chem.* **261**, 6643–6646.
9. Robinson, C. R. & Sauer, R. T. (1996) *Biochem.* **35**, 109–116.
10. O'Shea, E. K., Rutkowski, R. & Kim, P. S. (1992) *Cell* **68**, 699–708.
11. Pantoliano, M. W., Bird, R. E., Johnson, S., Asel, E. D., Dodd, S. W., Wood, J. F. & Hardman, K. D. (1991) *Biochem.* **30**, 10117–10125.
12. Mallender, W. D. & Voss, E. W., Jr. (1994) *J. Biol. Chem.* **269**, 199–206.
13. Rumbley, C. A., Denzin, L. K., Yantz, L., Tetin, S. Y. & Voss, E. W., Jr. (1993) *J. Biol. Chem.* **268**, 13667–13674.
14. Stemmer, W. P., Morris, S. K. & Wilson, B. S. (1993) *BioTechniques* **14**, 256–265.
15. Lieschke, G. J., Rao, P. K., Gately, M. K. & Mulligan, R. C. (1997) *Nat. Biotech.* **15**, 35–40.
16. Eustance, R. J. & Schleif, R. F. (1996) *J. Bacteriol.* **178**, 7025–7030.
17. Govindaraj, S. & Poulos, T. L. (1996) *Protein Sci.* **5**, 1389–1393.
18. Kortt, A. A., Lah, M., Oddie, G. W., Gruen, C. L., Burns, J. E., Pearce, L. A., Atwell, J. L., McCoy, A. J., Howlett, G. J., Metzger, D. W., et al. (1997) *Protein Eng.* **10**, 423–433.
19. Whitlow, M., Bell, B. A., Feng, S.-L., Filpula, D., Hardman, K. D., Hubert, S. L., Rollence, M. L., Wood, J. F., Schott, M. E., Milenic, D. E., et al. (1993) *Protein Eng.* **6**, 989–995.
20. Deonarain, M. P., Rowlinson-Busza, G., George, A. J. T. & Epenetos, A. A. (1997) *Protein Eng.* **10**, 89–98.
21. Tang, Y., Jiang, N., Parakh, C. & Hilvert, D. (1996) *J. Biol. Chem.* **271**, 15682–15686.
22. Newton, D. L., Xue, Y., Olson, K. A., Fett, J. W. & Rybak, S. M. (1996) *Biochem.* **35**, 545–553.
23. Huston, J. S., McCartney, J., Tai, M.-S., Mottola-Hartshorn, C., Jin, D., Warren, F., Keck, P. & Oppermann, H. (1993) *Int. Rev. Immunol.* **10**, 195–217.
24. Bowie, J. U. & Sauer, R. T. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2152–2156.
25. Milla, M. E., Brown, B. M. & Sauer, R. T. (1993) *Protein Sci.* **2**, 2198–2205.
26. Robinson, C. R. & Sauer, R. T. (1996) *Biochem.* **35**, 13878–13884.
27. Poljak, R. J. (1994) *Structure* **2**, 1121–1123.
28. Perisic, O., Webb, P. A., Holliger, P., Winter, G. & Williams, R. L. (1994) *Structure* **2**, 1217–1226.
29. Nagi, A. D. & Regan, L. (1997) *Fold. Des.* **2**, 67–75.
30. Chan, H. S. & Dill, K. A. (1988) *J. Chem. Phys.* **90**, 492–509.
31. Ramachandran, G. N. & Sasisekharan, V. (1968) *Adv. Protein Chem.* **23**, 283–437.
32. Breg, J. N., van Opheusden, J. H. J., Burgering, M. J. M., Roelens, R. & Kaptein, R. (1990) *Nature (London)* **346**, 586–589.
33. Raumann, B. E., Rould, M. A., Pabo, C. O. & Sauer, R. T. (1994) *Nature (London)* **367**, 754–757.
34. Kraulis, P. J. (1991) *J. Appl. Cryst.* **24**, 946–950.