Biochemistry

 Volume 35, Number 24 June 18, 1996

Accelerated Publications

Protein Stabilization by Removal of Unsatisfied Polar Groups: Computational Approaches and Experimental Tests†

Zachary S. Hendsch,‡ Thorlakur Jonsson,§ Robert T. Sauer,*,§ and Bruce Tidor*,‡

Departments of Chemistry and Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307

Received March 1, 1996; Revised Manuscript Received April 15, 1996⊗

ABSTRACT: The role of polar and charged side chains in partially buried protein environments has been probed in a variant of Arc repressor (MYL) in which hydrophobic interactions between Met31, Tyr36, and Leu40 replace the wild-type salt-bridge interactions between Arg31, Glu36, and Arg40. In the absence of this salt-bridge triad, three additional side chains were identified by continuum electrostatic calculations as incurring larger desolvation penalties during folding than were recovered in favorable electrostatic interactions in the folded state. These side chains (Asn29, Ser44, and Glu48) were mutated singly and collectively to alanine in the MYL background, and the thermodynamic stabilities of the resulting mutant proteins were found to be increased by 0.1 to 1.3 kcal/mol of dimer. All of the mutants displayed cooperative thermal melts and appeared to have well-packed hydrophobic cores by near-UV circular dichroism spectroscopy, indicating that conformational specificity is maintained. The Arc variant (MYL-NA29/SA44/EA48) in which the entire six-residue polar network is replaced by nonpolar groups is 5.1 kcal/mol of dimer more stable than wild type, indicating that the strategy of replacing buried or partially buried charged and polar side chains with hydrophobic residues can lead to substantial stabilization.

Interactions of charged and polar groups are important in the folding, binding, and biological activities of most macromolecules. The role of electrostatic forces, including solvation, in these processes can be probed by computational studies [for reviews see Honig and Nichols (1995) and references therein]. In the process of protein folding, for instance, calculations indicate that the magnitudes of electrostatic interactions are generally quite large but their net effects are frequently much smaller. This occurs because interactions of polar groups with solvent in the unfolded state are exchanged for interactions with other polar and charged groups in the folded protein. As a result, electrostatic interactions may stabilize or destabilize proteins depending on this balance of effects.

In previous work, continuum electrostatic calculations on 21 protein salt bridges showed that the electrostatic penalty for desolvation of the charged side chains upon protein folding was rarely recovered in favorable electrostatic interactions in the folded state (Hendsch & Tidor, 1994). These results suggested that replacing salt-bridge groups in proteins with hydrophobic groups of similar size and shape should generally lead to increased protein stability. Precisely this result was obtained in an experimental study of the Arc repressor homodimer, which contains two partially buried salt-bridge triads formed by the side chains of Arg31, Glu36, and Arg40 (see Figure 1). Following combinatorial mutagenesis of these codon positions, a number of hyperstable variants were recovered in which each of the salt-bridge residues was replaced by hydrophobic groups (Waldburger et al., 1995). One of these mutants, called MYL (Met31,

† Supported by NIH Grants AI-15706 and GM-47678 and a Science Partnership grant from MIT.

‡ Department of Chemistry.

§ Department of Biology.

⊗ Abstract published in *Advance ACS Abstracts,* May 15, 1996.

S0006-2960(96)00519-3 CCC: $12.00

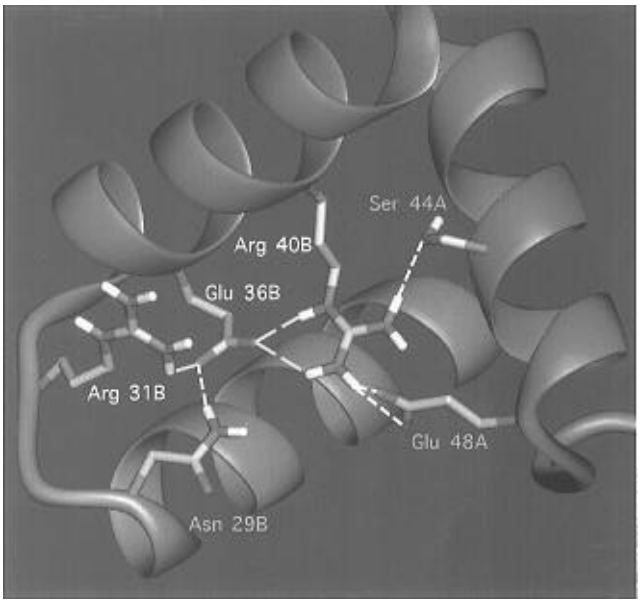

FIGURE 1: View of the Arg31−Glu36−Arg40 salt-bridge triad and adjoining residues in wild-type Arc. Note the Asn29−Glu36, Glu48−Arg40, and Ser44−Arg40 hydrogen bonds. Hydrogens were placed on the X-ray crystal structure using HBUILD (Brünger & Karplus, 1988).

Tyr36, and Leu40), was shown to be 3.9 kcal/mol of dimer more stable than wild type. In the crystal structure of this fully active mutant, the fold of the polypeptide main chain was found to be essentially wild type, with the chief structural changes being the substitution of hydrophobic interactions between the mutant side chains for the salt-bridge interactions between the wild-type side chains.

In the current work, we investigate the role of interactions between other residues in native Arc and the Arg31−Glu36−Arg40 salt-bridge triad. In particular, we use computational studies to identify polar and charged groups in Arc that appear to be orphaned by the absence of the salt-bridge triad in the MYL protein. We then show that the removal of these "locally unsatisfied" polar groups by mutagenesis leads to further stabilization of the MYL protein. Thus, the computational identification of unfavorable electrostatic interactions in native proteins and their subsequent elimination provides a reasonable strategy for achieving enhanced protein stability. Our results also underscore the remarkably strong context dependence of the structural information encoded in protein sequences.

METHODS

Identification of Candidate Residues. Continuum electrostatic calculations were used initially to identify Arc side chains with significant interactions with the Arg31−Glu36−Arg40 salt-bridge triad. The pairwise interaction ($\Delta\Delta G_{bridge}$) between each side chain in the protein and Arg31, Glu36, or Arg40 was calculated by charging one of the side chains and taking the sum of the charge times the potential at the other side chain as described previously (Hendsch & Tidor, 1994). Note that this is not simply the energy that would be calculated by Coulomb's law, since the presence of continuum solvent screens all interactions. Side chains with a favorable interaction energy of greater than 1 kcal/mol with any member of the salt-bridge triad were selected as candidates for further analysis.

Detailed Analysis. The electrostatic effect of each candidate side chain on the free energy of folding was calculated as the sum of its desolvation penalty ($\Delta\Delta G_{solv}$) and its electrostatic interactions in the folded protein ignoring Arg31, Glu36, and Arg40 ($\Delta\Delta G_{prot}$) (Hendsch & Tidor, 1994). For a given side chain, $\Delta\Delta G_{solv}$ was calculated as the difference in the reaction-field energy in folded- and unfolded-state models with charges present only on the selected side chain. $\Delta\Delta G_{prot}$ was calculated by charging just the selected side chain in the folded protein and summing the charge times the potential for all other atoms in the protein except those of the side chains of residues 31, 36, and 40 in each monomer of the Arc dimer. At the time of this work, the crystal structure of the MYL mutant had not been completed, and thus the electrostatic calculations were performed using the AB dimer (one of two dimers in the asymmetric unit) from the 1.8 Å resolution crystal structure of wild-type Arc (U. Obeysekare, C. Kissinger, L. J. Keefe, B. E. Raumann, R. T. Sauer and C. O. Pabo, unpublished). The effect of the unfolded state was incorporated through the desolvation energy of individual side chains. The unfolded-state model considered individual side chains to be fully solvent exposed.

Hydrogens were built into the crystal structure using the CHARMM HBUILD facility (Brünger & Karplus, 1988). The CHARMM (PARAM19) charge and radius parameters were used (Brooks et al., 1983) with a dielectric of 4 for the protein and 80 for solvent. A probe radius of 1.4 Å was used to define the surface. The linearized Poisson−Boltzmann equation was solved with an ionic strength of 145 mM and an ion exclusion layer of 2.0 Å using the program DelPhi (Gilson et al., 1988) as previously described (Hendsch & Tidor, 1994).

Mutagenesis and Purification. Arc variants were constructed by standard methods of cassette mutagenesis in the *arc-MYL-st11* gene of plasmid pSA700 (Milla et al., 1993; Brown et al., 1994). The sequences of mutant genes were verified by DNA sequencing. The isolation, purification, and properties of the MYL mutant have been described (Waldburger et al., 1995). Proteins were expressed in *Escherichia coli* strain X90, containing the appropriate pSA700 derivative. Each Arc variant contains an 11-residue C-terminal histidine-rich tail sequence (*st11*) to allow affinity purification. Proteins were purified to greater than 95% homogeneity as described (Milla et al., 1993) by chromatography on Ni^{2+}-NTA resin (Qiagen) and by ion-exchange chromatography on SP-Sephadex (Pharmacia).

Equilibrium Denaturation. Unless noted, a buffer containing 50 mM Tris-HCl buffer (pH 7.5) and 50 mM KCl was used for denaturation experiments. Guanidine hydrochloride denaturation at 25 °C was performed at a protein concentration of 2 μM (monomer equivalents), by monitoring the change in ellipticity at 222 nm on an Aviv-60DS circular dichroism (CD) spectropolarimeter. Thermal denaturation was performed at a protein concentration of 10 μM in buffer containing 2 M GuHCl. The addition of GuHCl allows the determination of the baseline ellipticity of the unfolded state of these hyperstable Arc variants. Denaturation data were fitted using nonlinear least squares procedures as previously described (Milla et al., 1993).

Near-UV Circular Dichroism. Because of the small molar ellipticity in the near-UV, CD spectra were determined at a protein concentration of 100 μM in a buffer containing 50

Table 1: Calculation of the Electrostatic Contributions of Asn29, Ser44, and Glu48 to Protein Stability in the Absence of Interactions with the Wild-Type Arg31−Glu36−Arg40 Salt-Bridge Triad[a]

	Asn29	Ser44	Glu48
$\Delta\Delta G_{solv}$	2.60	3.92	6.44
$\Delta\Delta G_{prot}$	0.10	0.56	−1.70
$\Delta\Delta G_{sum}$[b]	2.70	4.48	4.74

[a] All values are in kcal/mol of dimer. A positive value indicates an unfavorable electrostatic contribution to folding. $\Delta\Delta G_{solv}$ and $\Delta\Delta G_{prot}$ are calculated as described in the text. [b] $\Delta\Delta G_{sum} = \Delta\Delta G_{solv} + \Delta\Delta G_{prot}$ and is the predicted value for mutating the residue to a hydrophobic isostere.

mM Tris-HCl (pH 7.5) and 250 mM KCl. A set of five individual scans was averaged for each protein studied.

RESULTS

Electrostatic Calculations and Predictions. Three side chains in wild-type Arc were found to have interactions of 1 kcal/mol or more with one of the members of the Arg31−Glu36−Arg40 salt-bridge triad in the B monomer: Ser44A with Arg40B (−1.91 kcal/mol), Glu48A with Arg40B (−6.61 kcal/mol), and Asn29B with Glu36B (−3.21 kcal/mol). Symmetric interactions are expected to occur with the salt-bridge triad in the A monomer (i.e., Ser44B with Arg40A, but there is some missing electron density in this region of the A monomer in the X-ray structure) and thus the interaction energies are roughly twice as large for the Arc dimer. No groups were found to make strong electrostatic interactions with Arg31, possibly because it is the most solvent exposed of the triad and thus its interactions would be screened by solvent to the largest extent. As shown in Figure 1, the side chain of Asn29B hydrogen bonds to Glu36B, while the side chains of Ser44A and Glu48A hydrogen bond to Arg40B. Thus, in wild-type Arc, an extended hydrogen-bond/salt-bridge network is formed by the side chains of Asn29, Arg31, Glu36, Arg40, Ser44, and Glu48.

The results of further calculations (listed in Table 1) show that Asn29, Ser44, and Glu48 incur substantial desolvation penalties (2.6−6.4 kcal/mol of dimer) but make only modest interactions with the remainder of the protein (−1.7 to 0.5 kcal/mol of dimer). Thus, when electrostatic interactions with the residues of the wild-type salt-bridge triad are ignored, Asn29, Ser44, and Glu48 pay larger desolvation penalties than they recover in electrostatic interactions in the folded protein. On the basis of these calculations, Asn29, Ser44, and Glu48 should destabilize the native structure of the MYL protein, and thus changing these residues to alanine might be expected to increase the stability of the mutant proteins. This is an interesting prediction since the NA29, SA44, and EA48 mutations are known to *destabilize* otherwise wild-type Arc by 1.6−2.4 kcal/mol of dimer (Milla et al., 1994).

Experimental Tests. To determine the stability effects of the NA29, SA44, and EA48 mutations in the MYL background, genes containing each individual mutation and the triple substitution (NA29/SA44/EA48) were constructed, and the mutant proteins were purified. The equilibrium stabilities of these mutants were probed by guanidine hydrochloride denaturation at 25 °C and by temperature denaturation in a buffer containing 2 M GuHCl as shown in Figure 2 for the

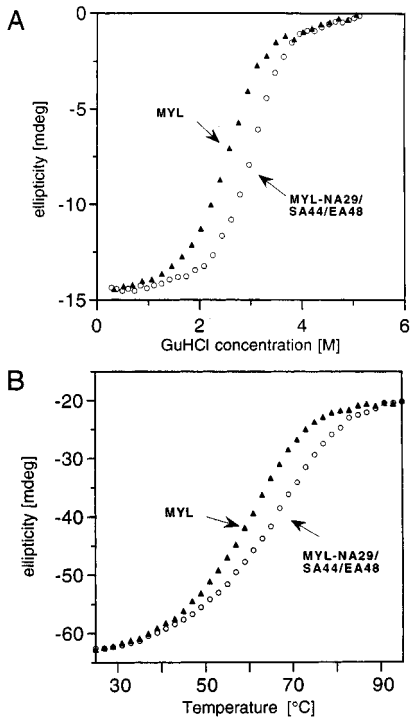

FIGURE 2: (A) Guanidine hydrochloride denaturation of the MYL and MYL-NA29/SA44/EA48 variants at a protein concentration of 2 μM in a buffer containing 50 mM Tris-HCl (pH 7.5) and 50 mM KCl monitored by changes in CD ellipticity at 222 nm. (B) Thermal denaturation of the MYL and MYL-NA29/SA44/EA48 variants at a protein concentration of 10 μM in a buffer containing 50 mM Tris-HCl (pH 7.5), 50 mM KCl, and 2 M GuHCl monitored by changes in CD ellipticity at 222 nm.

Table 2: Stability of MYL and Variants (50 mM Tris, pH 7.5, 50 mM KCl)

	ΔG_u[a] (kcal/mol)	$\Delta\Delta G_u$ (25 °C) (kcal/mol)	T_m[b] (°C)
MYL	14.28		56.8
MYL-EA48	14.41	0.13	58.4
MYL-NA29	14.94	0.66	59.3
MYL-SA44	15.26	0.98	64.3
MYL-NA29/SA44/EA48	15.60	1.32	66.0

[a] Free energy of dimer unfolding/dissociation at 25 °C, 1 M standard state, determined from fitting of experiments like those shown in Figure 2A. The value for *m*, the slope of ΔG_u vs [GuHCl], was fixed at −2.6 in the fitting procedure. Fitting the data independently, varying *m*, gave an average value for *m* of −2.59 ± 0.08 kcal/(mol of dimer·M). [b] Determined from experiments like those shown in Figure 2B. The value for ΔC_p was fixed at −1.1 in the fitting procedure. Independent curve fitting gave an average value for ΔC_p of −1.08 ± 0.07 kcal/(mol of dimer·K).

MYL and MYL-NA29/SA44/EA48 proteins. In both cases, the denaturation profiles are cooperative. The $\Delta\Delta G_u$ and T_m values calculated from the stability experiments are listed in Table 2. Consistent with the electrostatic calculations, each mutation is stabilizing in the MYL background. The MYL-SA44 variant is roughly 1 kcal/mol of dimer more stable than MYL while the MYL-NA29 and MYL-EA48 mutants show smaller increases in stability. The triple mutant is 1.3 kcal/mol of dimer more stable than the MYL parent. We note, however, that the stabilization afforded by the NA29, SA44, and EA48 mutations in the MYL background is significantly lower than would be predicted from the electrostatic calculations alone (see Discussion).

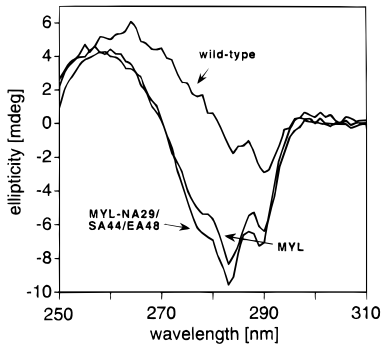

FIGURE 3: Near-UV circular dichroism spectra of wild-type Arc, MYL, and MYL-NA29/SA44/EA48 at a protein concentration of 100 μM in a buffer containing 50 mM Tris-HCl and 50 mM KCl. Five individual scans were averaged for each protein shown.

Close packing of aromatic groups in native protein cores gives rise to near-UV circular dichroism spectra which are dependent on the solvent and the proximity of other functional groups (Towell & Manning, 1994; Kahn, 1979; Strickland, 1974). By contrast, loosely packed molten globules do not give rise to significant CD spectra in the near-UV. As shown in Figure 3, wild-type Arc, MYL, and MYL-NA29/SA44/EA48 all display significant near-UV CD spectra, indicating that each protein has a tightly packed core. The MYL-NA29, MYL-SA44, and MYL-EA48 mutants also had near-UV spectra similar to those of MYL and the triple mutant (not shown). The increased negative signal in the 270−280 nm range and the corresponding decrease in signal in the 260−270 nm range for MYL and its variants, as compared with wild-type Arc, is probably caused by the presence of the partially buried Tyr36 side chain in these proteins. The overall similarities in the spectra of MYL and the MYL variants indicate that the alanine substitutions at positions 29, 44, and 48 do not appreciably perturb the packing of the protein.

DISCUSSION

Using electrostatic continuum calculations, we found that the side chains of Asn29, Ser44, and Glu48 interacted strongly and favorably with the 31−36−40 salt-bridge triad in wild-type Arc. In the absence of these interactions with the salt bridge, however, the electrostatic cost of desolvating these residues upon folding was not recovered in interactions with the remainder of the protein. This led to the prediction that substitution of alanine for Asn29, Ser44, and Glu48 would increase the stability of MYL, an Arc variant in which the 31−36−40 salt-bridge triad is replaced by hydrophobic residues. Experimentally, the three residues were mutated individually and in combination to alanine in the MYL background, and the resulting variants were found to display 0.1−1.3 kcal/mol of dimer increases in stability.

Although the MYL-NA29, MYL-SA44, and MYL-EA48 variants were indeed stabilized relative to MYL, the quantitative agreement between the stability changes expected by calculation and those observed experimentally was not very good. For example, stability increases of 2−5 kcal/mol of dimer for the single mutants were expected, but the largest observed stabilization was only about 1 kcal/mol of dimer. Several factors undoubtedly contribute to this discrepancy. First, the calculations that were done correspond to mutation of the wild-type residue to a hydrophobic isostere, whereas the experiments involve truncation of the wild-type side chain to the C_β position. Second, the calculations neglected any relaxation of the protein caused by the mutations, both from wild type to MYL and from MYL to the additional alanine mutants. That is, all calculations were carried out on wild-type Arc, which was the only structure available when this work was begun. Now that the MYL structure is available, we are in the process of examining whether structural changes allow Asn29, Ser44, and Glu48 to be solvated to a greater extent than in the native structure or to make improved electrostatic interactions elsewhere in the structure. An examination of the structure shows that Glu48 is closer to Arg50 to allow a stronger interaction between the two side chains in MYL, Asn29 rotates to hydrogen bond with a backbone carbonyl, and it is unclear how Ser44 is compensated, but it may direct its hydrogen toward Glu48. Finally, we note that the calculations make no attempt to include potential stability changes caused by altered packing, changes in buried hydrophobic surface, etc. In spite of these caveats, however, what makes this study both interesting and encouraging is that useful guidance was obtained from relatively quick calculations using the nonoptimal data available at the time.

The results presented here illustrate a simple strategy for stabilizing proteins: replace side chains that pay more in desolvation penalty than is recovered in intraprotein interactions. A related strategy of replacing a side chain in a strained rotamer by alanine resulted in enhanced stability in a study by Dao-Pin et al. (1990). While "electrostatically strained" side chains may be rare in wild-type proteins, replacement of tightly coupled electrostatic networks (such as the 31−36−40 salt-bridge triad in Arc) with hydrophobic groups that fill essentially the same volume not only can lead to more stable proteins (e.g., MYL) but also can serve as intermediates along a design pathway in which additional groups that form an extended polar network (see Figure 1) are also replaced. It has been suggested that buried polar interactions may help to define a unique folded state (Hendsch & Tidor, 1994), and some instances have been found in which polar-to-hydrophobic mutations lead to structural degeneracy (Harbury et al., 1993; Lumb & Kim, 1995). However, in both MYL (Waldburger et al., 1995) and the single and triple alanine mutants of MYL described here, there is no evidence for structural heterogeneity even though a large number of electrostatic interactions have been removed. Thus, replacing electrostatic interactions with hydrophobic interactions may be useful as a general method for increasing the stability of proteins.

In closing, we also wish to reiterate the importance of context in defining the relationship of protein sequence to structure and stability. In wild-type Arc, Asn29, Ser44, and Glu48 interact with the Arg31−Glu36−Arg40 salt-bridge triad as part of an extended electrostatic network and play roles in stabilizing the protein. In the wild-type background, substitution of single alanines for Asn29, Arg31, Glu36, Arg40, Ser44, or Glu48 leads to destabilization of the mutant protein by 1.6 to greater than 5 kcal/mol of dimer (Milla et al., 1994). Nevertheless, all six of these "important" polar and charged side chains can be replaced in combination by nonpolar residues with a net increase in stability of 5 kcal/mol of dimer. Thus, the roles of these residues in protein stability are not intrinsic but instead depend critically on the presence or absence of appropriate electrostatic partners. Once the salt-bridge triad is replaced in the MYL variant of

Arc, the structural roles and the information content of residues 29, 44, and 48 change radically.

ACKNOWLEDGMENT

We thank Carey Waldburger for helpful discussions and C. Kissinger and C. O. Pabo for the coordinates of the 1.8 Å resolution crystal structure of wild-type Arc.

REFERENCES

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.

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

Brünger, A. T., & Karplus, M. (1988) *Proteins: Struct., Funct., Genet. 4*, 148−156.

Dao-Pin, S., Baase, W. A., & Matthews, B. W. (1990) *Proteins: Struct., Funct., Genet. 7*, 198−204.

Gilson, M. K., Sharp, K. A., & Honig, B. (1988) *J. Comput. Chem. 9*, 327−335.

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

Hendsch, Z. S., & Tidor, B. (1994) *Protein Sci. 3*, 211−226.

Honig, B., & Nichols, A. (1995) *Science 268*, 1144−1149.

Kahn, P. (1979) *Methods Enzymol. 61*, 339−379.

Lumb, K. J., & Kim, P. S. (1995) *Biochemistry 34*, 8642−8648.

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

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

Strickland, E. H. (1974) *Crit. Rev. Biochem. 2*, 113−175.

Towell, J. F., & Manning, M. C. (1994) in *Analytical Applications of Circular Dichroism* (Purdie, N., & Brittain, H. G., Eds.) pp 175−205, Elsevier Science B. V., New York, NY.

Waldburger, C. D., Schildbach, J. F., & Sauer, R. T. (1995) *Nat. Struct. Biol. 2*, 122−128.

BI9605191