Tolerance of a protein to multiple polar-to-hydrophobic surface substitutions

MATTHEW H.J. CORDES AND ROBERT T. SAUER

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

(RECEIVED July 29, 1998; ACCEPTED October 16, 1998)

Abstract

Hydrophobic substitutions at solvent-exposed positions in two α -helical regions of the bacteriophage P22 Arc repressor were introduced by combinatorial mutagenesis. In helix A, hydrophobic residues were tolerated individually at each of the five positions examined, but multiple substitutions were poorly tolerated as shown by the finding that mutants with more than two additional hydrophobic residues were biologically inactive. Several inactive helix A variants were purified and found to have reduced thermal stability relative to wild-type Arc, with a rough correlation between the number of polar-to-hydrophobic substitutions and the magnitude of the stability defect. Quite different results were obtained in helix B, where variants with as many as five polar-to-hydrophobic substitutions were found to be biologically active and one variant with three hydrophobic substitutions had a t_m 6 °C higher than wild-type. By contrast, a helix A mutant with three similar polar-to-hydrophobic substitutions was 23 °C less stable than wild-type. Also, one set of three polar-to-hydrophobic substitutions in helix B was tolerated when introduced into the wild-type background but not when introduced into an equally active mutant having a nearly identical structure. Context effects occur both when comparing different regions of the same protein and when comparing the same region in two different homologues.

Keywords: aggregation; Arc repressor; combinatorial mutagenesis; denatured state; protein folding; thermal stability

Because it is energetically favorable for a polypeptide chain in aqueous solution to bury hydrophobic surface, most amino acid side chains that comprise the native protein core are apolar, whereas most that are solvent-exposed are polar (Miller et al., 1987). This correlation provides a useful general rule relating sequence to structure, but counterexamples are common. Surface positions, in particular, show considerable variation in polar character. Some solvent-exposed sites in natural proteins are occupied by hydrophobic side chains (see Miller et al., 1987) and mutagenesis studies show that many individual polar surface sites accept hydrophobic substitutions with little apparent consequence (Reidhaar-Olson & Sauer, 1988, 1990; Rennell et al., 1991; Suckow et al., 1996; Schwehm et al., 1998). Because proteins can tolerate some surface hydrophobics, it becomes pertinent to ask how many they can bear, whether some sites are more sensitive than others, and what consequences ensue when a protein has too many.

Placing apolar residues at solvent-exposed positions in a protein potentially may lead to changes in stability, solubility, or structure. If the protein can differentially bury the extra hydrophobic surface area in its native or denatured state, shifts in the folding/unfolding equilibrium would be expected (Pakula & Sauer, 1990). Alternatively, burial of the hydrophobics could occur in aggregated forms of either a folded or unfolded state, leading in the extreme case to

gross insolubility. Finally, the protein might be capable of adopting an alternative native structure in which the extra hydrophobic surface could be buried more efficiently (Bowie et al., 1990).

In the present study, groups of solvent-exposed polar residues in helix A or helix B of the bacteriophage P22 Arc repressor dimer were combinatorially mutated to allow substitution by hydrophobic amino acids. The effect of such mutations was found to depend strongly upon context. In helix A, an increase in the number of surface hydrophobics resulted in a decrease in folding stability, whereas in helix B a slight increase in stability was observed. The same set of polar-to-hydrophobic mutations in helix B also had different effects when introduced into the wild-type Arc background or into an active variant background with a higher density of hydrophobic residues. Thus, different regions on the surface of the same protein as well as the same region in different homologues vary in their tolerance to polar-to-hydrophobic mutations.

Results

Tolerance to polar-to-hydrophobic surface mutations in helix A

The structure of the Arc repressor is shown in Figure 1. Using cassette mutagenesis, five solvent-exposed positions in helix A (Glu17, Asp20, Lys24, Glu27, and Glu28, with side-chain fractional solvent accessibilities ranging from 0.50 to 0.87) were com-

Reprint requests to: Robert Sauer, 68-571, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139; e-mail: bobsauer@mit.edu.

Fig. 1. Structure of Arc repressor dimer (Raumann et al., 1994), showing secondary structure elements and helix A mutagenesis strategy. Symmetry-related side chains in helix A and N-terminal arm residues (1–7) are omitted for clarity.

binatorially substituted to allow four different amino acids, either the hydrophobic residues Val and Leu, or the polar residues Glu and Gln (Fig. 1). Sequencing of unselected clones indicated that this library contained approximately 57% hydrophobic substitutions and 43% polar substitutions overall. The combinatorial library was sorted into active and inactive classes using selection experiments for repressor activity in which cells expressing active Arc variants grow in the presence of streptomycin, and cells expressing inactive variants grow in the presence of chloramphenicol (Vershon et al., 1986; Bowie & Sauer, 1989b; Milla & Sauer, 1995). Comparisons between selected and unselected platings indicated that 10–15% of the mutants in the library were active. Figure 2A shows the sequences of 56 active variants and 29 inactive variants. In this library, variants are designated A- for helix A, followed by the one-letter code for the residues at positions 17, 20, 24, 27, and 28, respectively. For example, the mutant EV17/DV20/KE24/EL27/EQ28 is referred to as A-VVELQ.

Figure 2B shows distributions of the number of hydrophobic residues at the randomized positions in active and inactive sequences. For comparison, the expected distribution of the overall library is also shown. The active class shows a strong bias against hydrophobic substitutions, with only a small number of variants having as many as two and none having more than two polar-to-hydrophobic mutations. By contrast, the distribution of hydrophobic substitutions among the inactive variants approximates that of the overall library. This would be expected as inactive variants constitute 85–90% of the library. Table 1 lists the recovery of the

four possible amino acids at each randomized position in the active pool relative to the inactive pool. A value greater than 1.0 indicates over-representation in the active pool. Overall, at the set of positions, the activity assay selects strongly for glutamic acid (2.49), weakly for glutamine (1.06), and strongly against the hydrophobic residues leucine (0.33), and valine (0.30). Each of the four amino acids is allowed individually at each of the five positions, suggesting that the prohibition against more than two hydrophobics is a general effect rather than one specific to certain mutated positions. There are, however, clear positional biases. Leucine and valine, for instance, were recovered mainly at positions 24 and 27 in active sequences, somewhat less commonly at position 20, and rarely at positions 17 and 28.

Several factors could cause an inactive phenotype in the helix A library. Specific disruption of Arc-operator interactions is unlikely because the mutated side chains do not contact the DNA (Raumann et al., 1994) and mutants bearing single alanine substitutions at positions 17, 20, 24, 27, and 28 are biologically active. DNA binding could be reduced, however, if the mutations lowered steady-state expression, altered the native Arc structure or reduced the intracellular concentration of active Arc dimers by reducing protein stability, i.e., perturbing the equilibrium between native dimers and denatured monomers.

Steady-state expression assays showed that slightly more than half of the inactive variants were expressed at levels within 50% of the wild-type level, whereas the remainder showed a range of reduced expression levels. Seven of the inactive variants were purified and characterized in an attempt to identify the cause of the inactive phenotype. For comparison, two of the active variants were also purified and studied. Denaturation studies of each of these nine mutants (monitored by changes in circular dichroism) showed reduced stability compared with wild-type, with a rough correlation between the number of surface hydrophobics and the magnitude of the stability defect (Fig. 3; Table 2). Further, among mutants with two hydrophobic substitutions (2H), those that exhibited an active phenotype showed greater thermal stability than those that did not. The most stable of the 2H mutants contains valine and leucine at their most favored positions among the active sequences, 24 and 27, respectively (see Table 1). These findings suggest that a major factor leading to the inactive phenotypes of the helix A mutants in vivo is reduced levels of active dimers caused by a shift in the equilibrium toward denatured monomers. Because denatured monomers are also more susceptible to degradation, this could also explain reduced expression levels.

Table 1. Distribution of active and inactive sequences in helix A library according to position and residue^a

Residue type	Position					All
	17	20	24	27	28	
Leu	0.10	0.17	0.73	1.38	0.05	0.33
Val	0.11	0.24	0.73	0.30	0.06	0.30
Gln	2.59	1.14	1.04	0.69	0.78	1.06
Glu	1.71	3.83	1.75	2.85	4.14	2.49

^a Values in table indicate the probability of finding a particular residue at a particular position in active vs. inactive variants. Values are normalized for differences in total inactive and active sample sizes.

Fig. 2. A: Sequences of active and inactive variants from the helix A library, grouped by number of polar-to-hydrophobic substitutions, e.g., 0H, 1H. **B:** Distribution of helix A library classes by total number of hydrophobics at positions 17, 20, 24, 27, and 28. The overall distribution is simulated from the overall hydrophobic/polar composition of the library (57% H, 43% P). Isolates with the same amino acid sequence were generally found to have different DNA sequences and are therefore counted separately in determining distributions here and in Table 1. In the active variant distribution, a portion of the 2H bar is shaded in gray to indicate two variants (A-LQQVE and A-VQEVE) with partial activity.

We note that the thermal denaturation data shown in Figure 3 and Table 2 were collected under conditions where the melts were largely reversible. For several variants, particularly those with three or four hydrophobic substitutions, melts showed poor reversibility under standard conditions and denaturation occurred at lower temperature with increasing concentration, suggesting that the denatured state was prone to aggregation. For these mutants, we found that reversibility could be achieved by the addition of 22.5% glycerol. Under reversible melting conditions, most of the mutants showed a concentration dependence of stability similar to that of wild-type (i.e., increased stability at higher concentration), consistent with the expected equilibrium between native dimers and denatured monomers. In one case (A-VVELQ), however, the concentration dependence of stability was significantly reduced relative to wild-type (Table 2), suggesting that some oligomerization of denatured monomers may be occurring.

Hydrophobic residues on a protein surface could reduce stability if they were shielded from solvent more efficiently in the denatured state than in the native state (Pakula & Sauer, 1990; Bowler et al., 1993; Herrmann et al., 1995). To investigate this possibility for the helix A mutants, fluorescence spectra were taken of heat-denatured (75–80°C) or acid-denatured (pH 2.2) proteins. Relative to wild-type Arc, the thermally denatured mutants showed

Fig. 3. Representative thermal denaturation curves (monitored by circular dichroism) for helix A variants, illustrating decrease in thermal stability with increase in number of polar-to-hydrophobic substitutions. Denaturation is at 5 μ M protein, 50 mM Tris (pH 7.5), 250 mM KCl, 0.2 mM EDTA, 22.5% glycerol. For A-VVELQ and A-VLQVL, the reduced negative ellipticity of the low temperature baseline likely reflects a mixed population of denatured and native states due to low stability, rather than differences in native structure.

Table 2. Activity *in vivo* and thermal stability *in vitro* of selected helix A and helix B Arc mutants

Mutant sequence	Number of hydrophobic substitutions	Repressor activity in vivo	Thermal stability (relative to wt Arc) Δt_m (°C)	
			5 μ M	20 μ M
Helix A (17-20-24-27-28)				
A-EDKEE (wt)				
A-EEVLE ^a	2	Active	−4.2	−4.8
A-EELVE ^a	2	Active	−6.3	−7.0
A-EVLQE ^b	2	Inactive	−8.0	−9.0
A-ELEQL ^a	2	Inactive	−10.6	−10.5
A-EEVQV ^a	2	Inactive	−15.9	−14.1
A-VVELQ ^b	3	Inactive	−19.5	−24.5
A-QVLVQ ^b	3	Inactive	−23.0	−21.5
A-VLQVL ^b	4	Inactive	−26.4	—
A-LVELL	4	Inactive	—	—
Helix B (39-43-46-47)				
B-QEKK (wt)				
B-VLVQ ^b	3	Active	+5.7	—
B-LLVE	3	Active	—	—
B-VVVQ	3	Active	—	—
B-VVLE	3	Active	—	—
B-LVVL	4	Partial ^c	—	—
B-LLLV(S44I)	5	Partial ^c	—	—

^a t_m measured in 50 mM Tris, 250 mM KCl, 0.2 mM EDTA, pH 7.5.^b t_m measured in 50 mM Tris, 250 mM KCl, 22.5% glycerol, 0.2 mM EDTA, pH 7.5.^cIndicates growth on both streptomycin and chloramphenicol.

blue shifts of no more than 0.2 nm in the fluorescence emission center of mass but some of the pH-denatured mutants had blue-shifts as high as 7 nm (Table 3). These results show that the introduction of hydrophobic residues at surface positions in the native protein can lead to decreased solvation of the single tryptophan (Trp14) in some of the denatured mutant proteins. It is worth emphasizing, however, that some of the mutants do not show this behavior or only show significant blue shifts in the acid-denatured state.

Table 3. Shifts (nm) relative to wild-type in tryptophan fluorescence emission spectra (300–420 nm) of heat- and acid-denatured states of selected variants^a

Variant	Heat denatured ^b	Acid denatured ^c
A-EVLQE	0.0	−6.7
A-ELEQL	−0.2	−1.3
A-EEVQV	−0.1	−0.4
A-VVELQ	+0.1	0.0
A-VLQVL	−0.2	−0.3
A-LVELL	—	−2.6
B-VLVQ	−1.1	—

^aShifts are reported as differences in center of spectral mass (nm), as calculated in Materials and methods.^bMeasured at 75–85 °C under the buffer conditions used for thermal melts. Corrections were applied to some variants to account for small residual populations of native protein under the conditions (see Materials and methods).^cMeasured in 10 mM phosphoric acid, 250 mM KCl, pH 2.2.

The helix A variants do not appear to contain gross changes in native structure. Although the 3H mutants A-VVELQ and A-QVLVQ and the 4H mutant A-VLQVL did not show fully wild-type-like CD or fluorescence spectra, these differences are consistent with the low stabilities of these mutants, resulting in the persistence of a population of denatured molecules under all conditions studied. Further, even mutants that were inactive *in vivo* showed some binding of wild-type operator DNA *in vitro*, and reductions in binding activity compared to wild-type Arc paralleled decreases in stability (data not shown).

Tolerance to surface mutations in helix B

As discussed above, mutants with three or more surface hydrophobic substitutions in helix A were inactive *in vivo* and showed poor stability and a propensity to aggregate *in vitro*. To determine the effects of similar mutations in helix B, a limited combinatorial mutagenesis was performed in which three solvent-exposed positions (Gln39, Glu43, and Lys46, with side-chain fractional solvent accessibilities from 0.53–0.66) were substituted with hydrophobics (Val or Leu), and a fourth position (Lys47, with side-chain fractional solvent accessibility of 0.73) with hydrophobic and polar residues (Val, Leu, Glu, or Gln). Mutants from this library were thus designed to contain either three or four hydrophobic substitutions. As with the helix A library, side chains at these four positions are not directly required for Arc repressor function and do not contact operator DNA (Raumann et al., 1994).

Six mutants with three, four, or five hydrophobic substitutions were isolated from the helix B library (Table 2). The mutant with five changes contained a mutation of Ser44 to Ile even though this

position was not mutagenized. These variants are designated B- for helix B, followed by the one-letter code for the residues at positions 39, 43, 46, and 47, respectively. Activity assays revealed that the four variants with three hydrophobics were fully active, and those with four or five hydrophobic substitutions were partially active. These results are strikingly different than those for helix A, in which no mutants with more than two hydrophobic substitutions showed even partial activity.

The B-VLVQ mutant was purified and characterized *in vitro*. Efforts to characterize other helix B mutants were hampered by poor solubility. Similar to the 3H mutants in helix A, the thermal melting behavior of B-VLVQ showed poor reversibility under standard conditions, but was reversible in the presence of 22.5% glycerol. In the latter solvent, B-VLVQ showed increased thermal stability relative to wild-type Arc, with a 6 °C increase in melting temperature. Fluorescence and circular dichroism (CD) spectra of B-VLVQ show evidence for increases in hydrophobic burial and structure in the thermally denatured protein. At high temperature (80–85 °C), B-VLVQ shows a 1.1 nm blue shift in the center of mass of its fluorescence emission spectrum relative to thermally denatured wild-type Arc, an effect greater than those seen for helix A mutants. The unfolded baseline of the thermal melt of B-VLVQ by CD at 222 nm also shows slightly more signal than that of wild-type Arc (see Fig. 4), whereas no such effect is apparent for helix A variants (see Fig. 3).

The helix B substitutions in B-VLVQ result in a 6 °C increase in stability, whereas a similar set of substitutions in the helix A

mutant, A-QVLVQ, causes a 23 °C decrease in stability. Because the VLVQ substitutions are introduced with identical spacings in α -helical contexts in both mutants, different sequence or tertiary contexts must account for their markedly different stability effects. To investigate the importance of local sequence in determining the effect of the context, we introduced the B-VLVQ sequence substitutions into the Arc-MYL background (Waldburger et al., 1995). Arc-MYL is an active variant with the same structure as wild-type Arc but with Arg31, Glu36, and Arg40, which form a partially buried tripartite salt bridge, replaced by the hydrophobic residues Met31, Tyr36, and Leu40. Surface substitutions in helix B of wild-type Arc and Arc-MYL therefore occur in the same secondary and tertiary context but in different local sequence contexts. The MYL/B-VLVQ mutant was inactive *in vivo*, in contrast to the MYL and B-VLVQ parent molecules, which are fully active. Figure 4 shows that the CD signal of MYL/B-VLVQ at 222 nm falls between those of native and thermally denatured wild-type, MYL, or B-VLVQ, and does not change with increasing temperature. This suggests strongly that MYL/B-VLVQ does not adopt the native Arc fold. Two other results also support a non-native structure for this mutant. First, at 25 °C, the tryptophan fluorescence spectrum of MYL/B-VLVQ (344 nm) has a higher center of mass than that of MYL (340 nm), which indicates reduced tryptophan burial in the MYL/B-VLVQ mutant. Second, gel filtration of MYL/B-VLVQ at 25 °C revealed a broad distribution of high molecular weight aggregated species.

Fig. 4. Thermal denaturation curves for B-VLVQ mutants in two different sequence backgrounds compared to parent molecules: (A) wild-type Arc background and (B) Arc-MYL background. Sequences of parent and mutant molecules are shown at bottom.

The Arc-MYL background is much less tolerant of the B-VLVQ sequence substitutions as compared to the wild-type Arc background. To investigate whether the effect of even single polar-to-hydrophobic mutations shows such a dependence on local sequence context, the Arc/Q39L and MYL/Q39L mutants were constructed, purified, and characterized. As shown in Table 4, the Q39L mutation preserves *in vivo* activity and confers a similar, slight increase in thermal stability in both the Arc and Arc-MYL contexts. While these results seem to indicate that Arc-MYL tolerates single polar-to-hydrophobic surface substitutions as well as wild-type Arc, it should be noted that the MYL and MYL/Q39L thermal melts were performed in 2 M GuHCl (Hendsch et al., 1996). In the absence of GuHCl, MYL/Q39L is prone to massive aggregation upon heating, whereas Arc/Q39L shows reversible, two-state melting. Thus, despite the absence of a discernible contextual effect on stability under reversible conditions, a single polar-to-hydrophobic mutation leads to aggregation behavior in Arc-MYL but not in wild-type Arc.

Discussion

A number of mutagenesis and evolutionary studies (Perutz et al., 1965; Reidhaar-Olson & Sauer, 1988, 1990; Rennell et al., 1991; Suckow et al., 1996; Schwehm et al., 1998) have shown that most surface positions in proteins tolerate amino acid substitutions even when these involve changes from polar to hydrophobic residues. These same studies, however, also show that certain surface positions are more tolerant of hydrophobic residues than others. For example, activity selections on λ repressor mutants show that 70% of solvent-exposed side chains accepted hydrophobic replacements, although at some of these positions bulky hydrophobics were prohibited (Reidhaar-Olson & Sauer, 1988, 1990). Similarly, single phenylalanine substitutions at 27 solvent-exposed positions in staphylococcal nuclease led to an average decrease in stability of 0.5 kcal/mol, but some of these substitutions actually stabilized the protein whereas others were destabilizing by as much as 2.0 kcal/mol (Schwehm et al., 1998).

The combinatorial mutagenesis results reported here for nine polar surface residues in Arc confirm both the general tolerance of surface residues to substitution and also the variability in the sensitivity of sites to hydrophobic replacements. Activity assays showed that each of the nine mutagenized surface residues tolerate the hydrophobic residues leucine and valine. In active variants from the helix A library, however, these amino acids are found much more commonly at positions 24 and 27 than at positions 17, 20, or 28. Moreover, although polar-to-hydrophobic substitutions are well-tolerated individually, our results reveal clear albeit context-dependent limits

on how many may be introduced at once. In helix A of Arc, only a small number of variants with as many as two hydrophobic mutations, and none with more than two, were active. Purified helix A variants with three or four hydrophobic substitutions had reduced thermal stability and tended to form denatured aggregates, with both factors probably contributing to loss of biological activity. These findings suggest that the predominance of polar residues at surface positions in proteins may have its origin primarily in the requirement that the density of hydrophobic surface residues be limited. Our results also show that some regions of a protein can tolerate more surface hydrophobics than others. Thus, helix B variants with as many as five hydrophobic substitutions had biological activity. Finally, we observed that the same mutations were tolerated very differently in two equally active protein backgrounds. The wild-type Arc, Arc-MYL, and Arc/B-VLVQ proteins were all active and well folded, whereas the Arc-MYL/B-VLVQ variant was inactive and formed non-native structures.

What causes the decreasing stability with an increasing numbers of hydrophobic substitutions in the helix A mutants and why is the situation different for the helix B mutants? Both the magnitude of the stability changes and their dependence on context are poorly predicted by the thermal stabilities of mutants with single alanine substitutions, suggesting that these results cannot be explained by simple preferences for certain wild-type side chains. For example, adding the Δt_m 's (which are linearly correlated with $\Delta\Delta G$ values for Arc mutants) of the individual alanine substitutions at the five mutated positions in helix A predicts a stability decrement of 9 °C and the same analysis of the four helix B positions predicts an 11 °C destabilization (Milla et al., 1994). Nevertheless, helix A mutants with three hydrophobic substitutions show stability decreases of approximately 20 °C, whereas the 3H helix B mutant B-VLVQ shows a 6 °C stability increase. Intrinsic secondary-structure propensities are also unlikely to contribute significantly to the stability effects seen here. Among residues used in the helix A mutagenesis, leucine is the most helix-forming but was no more commonly found in active variants than valine, which is a strong helix breaker. Further, the helix A mutant A-QVLVQ and helix B mutant B-VLVQ have equal numbers of leucines and valines but differed markedly in stability.

The destabilization observed in the helix A mutants might, in part, reflect the influence of charged-to-neutral substitutions on local or global electrostatic interactions in the native Arc dimer, which normally carries a net charge of +8 at neutral pH. Placing neutral side chains at the five substituted residues in helix A increases the net charge of the mutant dimers to +14, whereas the four polar-to-hydrophobic substitutions in helix B reduce the net charge of the mutant dimers to +6. Indeed, a destabilizing effect of extra net positive charge is consistent with the fact that wild-type Arc is stabilized by increasing salt and destabilized by decreasing pH (Bowie & Sauer, 1989b). Charge effects are also consistent with our finding that the neutral glutamine side chain occurs less frequently in active variants from the helix A library than the isosteric but negatively charged glutamic acid side chain. Electrostatic effects cannot, however, be the whole story, as leucine and valine are tolerated less well than glutamine in the helix A experiments, even though all three are neutral. In addition, any straightforward effect of electrostatic interactions should be reflected in the stabilities of single alanine mutants but, as noted above, this is not the case.

Effects on the denatured state may be as important as effects on the native state in explaining the effects of polar-to-hydrophobic

Table 4. Q39L mutant activity and thermal denaturation data

Variant	Repressor activity <i>in vivo</i>	t_m (°C)
Wild-type Arc	Active	58.3 ^a
Arc/Q39L	Active	64.0 ^a
MYL	Active	57.2 ^b
MYL/Q39L	Active	62.2 ^b

^a At 10 μ M protein, 50 mM Tris, 250 mM KCl, 0.2 mM EDTA, pH 7.5.

^b At 10 μ M protein, 50 mM Tris, 50 mM KCl, 2 M GuHCl, pH 7.5.

surface substitutions on protein stability. Experimental and theoretical studies suggest that the denatured states of proteins are not open random-coil configurations but, in fact, are compact ensembles with some residual structure (Dill & Shortle, 1991). Thermally denatured Arc, for example, shows significant CD ellipticity and forms a specific dimer at sufficiently high concentrations (Robinson et al., 1997). In principle, a hydrophobic side chain introduced at certain solvent-exposed positions in a protein might be buried more effectively in the ensemble of denatured states than in the native protein. Such behavior, termed a reverse hydrophobic effect, would be expected to shift the folding equilibrium toward denaturation (Pakula & Sauer, 1990; Bowler et al., 1993; Herrmann et al., 1995). Some of the denatured mutant proteins studied here did show evidence of increased hydrophobic burial as monitored by blueshifts in tryptophan fluorescence emission spectra, but there was no consistent trend or correlation between the destabilization and the magnitude of these blueshifts. One obvious problem in interpreting such experiments is that fluorescence only probes the environment of the single tryptophan residue at position 14 in Arc. Hence, the side chains introduced by certain polar-to-hydrophobic substitutions might be more highly buried in the denatured state without the mutant protein displaying increased tryptophan burial.

In this study, Arc and the active Arc variant MYL were found to exhibit dramatically different responses to hydrophobic substitutions in helix B. The Arc/Q39L and Arc/B-VLVQ mutants, which contain one and three additional hydrophobic residues, respectively, were both 6°C more thermally stable than the parent molecules and were both active *in vivo*. By contrast, the MYL/Q39L and MYL/B-VLVQ mutants, which contain four and six additional hydrophobics, respectively, behave quite differently. MYL/Q39L is active and, like Arc/Q39L, more stable than its parent. MYL/B-VLVQ is inactive and does not fold into an Arc-like native state *in vitro*. Despite the stabilization observed with hydrophobic substitutions in helix B, it appears that a threshold exists of five or six additional hydrophobics relative to wild-type Arc, beyond which catastrophic aggregation/misfolding with concomitant loss of biological activity occur. Whether this is a result of too many hydrophobic residues in the sequence as a whole, or simply too high a local density of hydrophobics, is not clear. It seems reasonable that very hydrophobic local sequences could nucleate denatured state aggregation, and would thus be unfavorable. Indeed, surveys of the protein database show that pentapeptide sequences containing only hydrophobic residues occur far less frequently than would be statistically expected based on their single residue frequencies (Vazquez et al., 1993).

Variations in local hydrophobic sequence density could help to explain certain positional effects observed within helix A. For instance, with the exception of position 28, the positions in the N-terminal half of the helix (17 and 20) are more sensitive to mutation than those in the C-terminal half (24 and 27). The N-terminal half of helix A in wild-type is also much more dense in hydrophobic residues (see Fig. 1). Further, among 2H mutants, the only one studied that exhibited an irreversible thermal melt under standard conditions was A-EVLQE, which contains a stretch from 18–26 in which 8 of 9 residues are aliphatic. None of the other 2H mutants studied contains a stretch of such length and high hydrophobic density.

In conclusion, multiple polar-to-hydrophobic mutations in the two α -helices of Arc repressor lead to significant, context-dependent effects on the stability and aggregation properties of the mutant

proteins. No single factor adequately explains why multiple polar-to-hydrophobic substitutions are destabilizing in the helix A context but stabilizing in the helix B context, suggesting that a variety of physical factors contribute to the observed effects. We noted that changes in native structure might also be a potential consequence of hydrophobic mutations on a protein's surface. While no large effects of this kind were found in this study, we have recently found that polar-to-hydrophobic mutations in the β -sheet of Arc do, in fact, promote dramatic changes in local secondary and tertiary structure (M.H.J. Cordes, N.P. Walsh, C.J. McKnight, & R.T. Sauer, *in prep.*).

Materials and methods

Mutagenesis and selection

Codons for Arc residues 17, 20, 24, 27, and 28 in the helix A library were mutagenized by synthesis of the sense strand of a SacII/BglIII cassette, encoding Arc residues 15–32 (Bowie & Sauer, 1989a) with C and G at the first codon position, A and T at the second position, and A and G at the third position. This combination of bases results in codons encoding Gln, Glu, Leu, and Val. The 3' end of the sense strand contained a three-base extension designed to give a 12-base self-annealing sequence, which allowed enzymatic second-strand synthesis using Sequenase (United States Biochemical, Cleveland, Ohio). Digestion of the double stranded fragment with SacII and BglIII then yielded the desired cassette. The 5' end of the sense strand also contained a five base extension beyond the SacII sequence to ensure efficient restriction enzyme digestion. Cassettes were ligated to the large SacII/BglIII backbone fragment of pSA700, a plasmid that encodes Arc with the st11 C-terminal extension HHH-HHHKNQHE (Milla et al., 1993).

Mutagenesis of helix B was performed by randomizing codons 39, 43, 46, and 47 within a 60-base sense strand spanning the EcoRI site at codons 33–35 and extending to the codon for Ile51. The same randomization scheme was used at these codons as in helix A, except that for 39, 43, and 46 only T was allowed at the second codon position, so that only the hydrophobics Leu and Val were permitted at these positions. This oligonucleotide was annealed to an overlapping 64-base antisense strand, which extended just beyond the ClaI site at the 3' end of the open reading frame (Bowie & Sauer, 1989a). Enzymatic second-strand synthesis and restriction enzyme digestion then yielded an EcoRI/ClaI cassette, which was ligated to the large EcoRI/ClaI backbone fragment of pSA700. The mutant B-VLVQ/MYL was synthesized by analogous methods, using the large EcoRI/ClaI backbone fragment of pSA701, a plasmid that encodes Arc-MYL with the st11 C-terminal extension. The mutants Arc/Q39L and MYL/Q39L were synthesized by PCR mutagenesis of pSA700 or pSA701 using a mutagenic 3' primer with a single A to T base change at the second position of codon 39.

Ligation mixtures were transformed into *Escherichia coli* strain UA2F cells by electroporation, and chloramphenicol and streptomycin selection experiments performed largely as described (Milla & Sauer, 1995) but without the addition of isopropyl- β -thiogalactoside (IPTG).

Protein expression and purification

The steady-state expression levels of mutant proteins in strain UA2F/pSA700 were assayed by SDS-polyacrylamide gel electro-

phoresis of crude cell lysates essentially as described (Milla et al., 1993). Wild-type Arc-st11 and variants were purified from strains UA2F/pSA700 or X90/pSA700 by chromatography on Ni²⁺-NTA agarose (Qiagen, Inc., Valencia, California) and SP-sephadex (Pharmacia, Uppsala, Sweden) (Milla et al., 1993). MYL variants were purified on Ni²⁺-NTA agarose using 10 mM imidazole in sample load and wash steps. Protein concentrations in monomer equivalents per liter were determined using an extinction coefficient at 280 nm of 6,756 cm⁻¹ (Brown et al., 1990).

Biophysical methods

Unless otherwise indicated, all experiments on variants in the wild-type background were performed in a standard buffer containing 50 mM Tris-HCl (pH 7.5), 250 mM KCl, and 0.2 mM EDTA, or in the same buffer containing 22.5% glycerol. For MYL-based variants, experiments were performed in 50 mM Tris-HCl (pH 7.5), 50 mM KCl, 2 M GuHCl (Hendsch et al., 1996).

Circular dichroism spectra and melts were recorded on an AVIV Model 60DS spectrometer. For thermal melts, Arc (5–20 μM) was heated in 1 °C steps, the sample was allowed to equilibrate for 1 min at each temperature, and the ellipticity at 222 nm (5 μM) or 230 nm (20 μM) was averaged for 30 s. Most of the variants studied showed greater than 80% reversibility at 20 μM, though several showed only 70% reversibility. In such cases, t_m 's obtained from reverse melts agreed with forward melt values to within 1.5 °C. Melts were fit to a two-state model in which two denatured monomers are in equilibrium with the native dimer. Values of t_m , ΔH_u , and the slopes and intercepts of the native and denatured baselines were obtained from the fitting procedure, assuming a ΔC_p value of 1.31 kcal/mol (Milla et al., 1994). For MYL-based variants, melts were obtained in 50 mM Tris, 50 mM KCl, 2 M guanidine hydrochloride. Gel filtration of MYL/B-VLVQ was performed using a Superose 12 column.

Tryptophan fluorescence spectra were recorded on a Hitachi F-4500 fluorescence spectrophotometer with a thermostated cell holder heated by a Neslab RTE-220 heating bath. Spectra were obtained from 300 to 420 nm with excitation of the tryptophan chromophore at 280 nm, and the center of spectral mass $\langle \nu_p \rangle$ calculated from the equation:

$$\langle \nu_p \rangle = \frac{\sum_i F_i \nu_i}{\sum_i F_i} \quad (1)$$

where F_i is the fluorescence emitted at wavenumber ν_i , and the summation is carried out over the full range of the spectrum. Equation 1 yields center-of-mass values in frequency units (wavenumbers). Values reported in Table 4 are converted to wavelengths.

Low pH fluorescence spectra were measured at 25 °C in 10 mM phosphoric acid, 250 mM KCl, pH 2.2. High-temperature fluorescence spectra were measured at 75–85 °C in either standard buffer or standard buffer with 22.5% glycerol. The cell temperature was calibrated with a digital probe thermometer attached to the cell holder. Some variants contained a small residual population (0.5–1.5%) of native Arc at these temperatures, as determined by fitting of CD melts. In these cases, centers-of-mass were corrected to the denatured baseline by applying a fraction of the difference between

native and denatured values as measured from a center-of-mass fluorescence melt of wild-type Arc.

Acknowledgments

This research was supported by NIH grant AI-15706 and by a Helen Hay Whitney postdoctoral fellowship (M.H.J.C.).

References

- Bowie JU, Reidhaar-Olson JF, Lim WA, Sauer RT. 1990. Deciphering the message in protein sequences: Tolerance to amino acid substitutions. *Science* 247:1306–1310.
- Bowie JU, Sauer RT. 1989a. Identifying determinants of folding and activity for a protein of unknown structure. *Proc Natl Acad Sci USA* 86:2152–2156.
- Bowie JU, Sauer RT. 1989b. Equilibrium dissociation and unfolding of the Arc repressor dimer. *Biochemistry* 28:7139–7143.
- Bowler BE, May K, Zaragoza T, York P, Dong A, Caughey WS. 1993. Destabilizing effects of replacing a surface lysine of cytochrome *c* with aromatic amino acids: Implications for the denatured state. *Biochemistry* 32:183–190.
- Brown BM, Bowie JU, Sauer RT. 1990. Arc repressor is tetrameric when bound to operator DNA. *Biochemistry* 29:11189–11195.
- Dill KA, Shortle D. 1991. Denatured states of proteins. *Annu Rev Biochem* 60:795–825.
- Hendsch ZS, Jonsson T, Sauer RT, Tidor B. 1996. Protein stabilization by removal of unsatisfied polar groups: Computational approaches and experimental tests. *Biochemistry* 35:7621–7625.
- Herrmann L, Bowler BE, Dong A, Caughey WS. 1995. The effects of hydrophilic to hydrophobic surface mutations on the denatured state of iso-1-cytochrome *c*: Investigation of aliphatic residues. *Biochemistry* 34:3040–3047.
- Milla ME, Brown BM, Sauer RT. 1993. P22 Arc repressor: Enhanced expression of unstable mutants by addition of polar C-terminal sequences. *Protein Sci* 2:2198–2205.
- Milla ME, Brown BM, Sauer RT. 1994. Protein stability effects of a complete set of alanine substitutions in Arc repressor. *Nat Struct Biol* 1:518–523.
- Milla ME, Sauer RT. 1995. Critical side chain interactions at a subunit interface in the Arc repressor dimer. *Biochemistry* 34:3344–3351.
- Miller S, Janin J, Lesk AM, Chothia C. 1987. Interior and surface of monomeric proteins. *J Mol Biol* 196:641–656.
- Pakula AA, Sauer RT. 1990. Reverse hydrophobic effects relieved by amino acid substitutions at a protein surface. *Nature* 344:363–364.
- Perutz MF, Kendrew JC, Watson HC. 1965. Structure and function of haemoglobin. II. Some relations between polypeptide chain configuration and amino acid sequence. *J Mol Biol* 13:669–678.
- Raumann BE, Rould MA, Pabo CO, Sauer RT. 1994. DNA recognition by β -sheets in the Arc repressor-operator crystal structure. *Nature* 367:754–757.
- Reidhaar-Olson JF, Sauer RT. 1988. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science* 241:53–57.
- Reidhaar-Olson JF, Sauer RT. 1990. Functionally acceptable substitutions in two α -helical regions of λ repressor. *Proteins Struct Funct Genet* 7:306–316.
- Rennell D, Bouvier SE, Hardy LW, Poteete AR. 1991. Systematic mutation of bacteriophage T4 lysozyme. *J Mol Biol* 222:67–88.
- Robinson CR, Rentzeperis D, Silva JL, Sauer RT. 1997. Formation of a denatured dimer limits the thermal stability of Arc repressor. *J Mol Biol* 273:692–700.
- Schwehm JM, Kristyanne ES, Biggers CC, Stites WE. 1998. Stability effects of increasing the hydrophobicity of solvent-exposed side chains in staphylococcal nuclease. *Biochemistry* 37:6939–6948.
- Suckow J, Markiewicz P, Kleina LG, Miller J, Kisters-Woike B, Muller-Hill B. 1996. Genetic studies of the lac repressor. XV: 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure. *J Mol Biol* 261:509–523.
- Vazquez S, Thomas C, Lew RA, Humphreys RE. 1993. Favored and suppressed patterns of hydrophobic and nonhydrophobic amino acids in protein sequences. *Proc Natl Acad Sci USA* 90:9100–9104.
- Vershon AK, Bowie JU, Karplus TM, Sauer RT. 1986. Isolation and analysis of Arc repressor mutants: Evidence for an unusual mechanism of DNA binding. *Proteins Struct Funct Genet* 1:302–311.
- Waldburger CD, Schildbach JF, Sauer RT. 1995. Are buried salt bridges important for protein stability and conformational specificity? *Nat Struct Biol* 2:122–128.