Published in final edited form as:

Proteins. 2019 January ; 87(1): 23–33. doi:10.1002/prot.25613.

Increased sequence hydrophobicity reduces conformational specificity: a mutational case study of the Arc repressor protein

Katie L. Stewart^{*,†}, Deepali Rathore¹, Eric D. Dodds¹, and Matthew H. J. Cordes^{*,‡}

^{*}Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ 85721-0088 USA

¹Department of Chemistry, University of Nebraska, Lincoln, NE 68588-0304 USA

[†]present address: Department of Physics, Emory University, Atlanta, GA 30322-2430 USA

Abstract

The amino-acid sequences of soluble, globular proteins must have hydrophobic residues to form a stable core, but excess sequence hydrophobicity can lead to loss of native state conformational specificity and aggregation. Previous studies of polar-to-hydrophobic mutations in the β -sheet of the Arc repressor dimer showed that a single substitution at position 11 (N11L) leads to population of an alternate dimeric fold in which the β -sheet is replaced by helix. Two additional hydrophobic mutations at positions 9 and 13 (Q9V and R13V) lead to population of a differently folded octamer along with both dimeric folds. Here we conduct a comprehensive study of the sequence determinants of this progressive loss of fold specificity. We find that the alternate dimer fold specifically results from the N11L substitution and is not promoted by other hydrophobic substitutions in the β -sheet. We also find that three highly hydrophobic substitutions at positions 9, 11 and 13 are necessary and sufficient for oligomer formation, but the oligomer size depends on the identity of the hydrophobic residue in question. The hydrophobic substitutions increase thermal stability, illustrating how increased hydrophobicity can increase folding stability even as it degrades conformational specificity. The oligomeric variants are predicted to be aggregation-prone but may be hindered from doing so by proline residues that flank the β -sheet region. Loss of conformational specificity due to increased hydrophobicity can manifest itself at any level of structure, depending upon the specific mutations and the context in which they occur.

Keywords

Sequence hydrophobicity; protein folding; sequence-structure relationship; protein conformation; structural degeneracy; conformational specificity

INTRODUCTION

One of the most remarkable feats performed by soluble, globular proteins is folding to a stable, specific, unique three-dimensional structure, encoded by the amino-acid sequence. The sequence determinants of thermodynamic stability--the ability to fold to the native structure as opposed to remaining unfolded--are relatively well understood. The sequence

[‡]To whom correspondence may be addressed. phone: (520) 626-1175, fax: (520) 621-9288, cordes@email.arizona.edu.

determinants of conformational specificity--the ability to specify one folded structure versus an alternate structure--are less well understood¹. In protein design, stability requires a mostly positive design approach to optimize favorable interactions, while specificity also requires negative design to disfavor alternative conformations². Thus the two sets of determinants cannot be the same, and since the set of alternative conformations is rarely fully known, specificity is generally more challenging than stability to design and to understand³.

Sequence hydrophobicity is critical for folding stability but can be counterproductive for conformational specificity. The primary thermodynamic drive for protein folding is formation of a hydrophobic interior⁴⁻⁶, and proteins with very low overall sequence hydrophobicity are generally unfolded under native conditions⁷. On the other hand, an excess of stabilizing hydrophobic interactions increases the likelihood of alternative low-energy folded conformations in addition to the native state⁸. Simple computational models of proteins suggest that sequences with very high hydrophobicity are less likely to have a single most stable folded state^{9,10}. Mutations that increase hydrophobicity lead to loss of native state specificity in some cases, which can manifest as increased dynamic behavior and loss of specific tertiary or quaternary interactions¹¹⁻¹³, aggregation of the native state¹⁴, or more severe changes such as switches in folding topology¹⁵ and/or aggregation of nonnative, misfolded states¹⁶⁻¹⁸.

Mutational studies of Arc repressor have clearly demonstrated that increases in sequence hydrophobicity can progressively decrease structural specificity and thereby increase native state degeneracy. Arc repressor is a ribbon-helix-helix DNA-binding protein that folds as an intertwined homodimer (Figure 1). A single polar-to-hydrophobic substitution in the Arc repressor β -sheet (Arc-N11L) leads to population of a second fold in which the β -sheet is replaced by a helical structure (Figure 1A)¹⁵. Comparison of the dimer structures (Figure 1C) provides a simple rationale: position 11 is solvent-exposed in the wild-type structure, but buried in the core of the alternate fold¹⁹. The Q9V/N11L/R13V (called S-VLV) variant, which has two additional polar-to-hydrophobic mutations in the same β -sheet, forms an octamer with about half the α -helical content of wild-type Arc²⁰. S-VLV also retains the ability to form both dimer structures of Arc-N11L, demonstrating three distinct folded structures. The structure of the S-VLV octamer is unknown, but it seems likely that nonpolar side chains at positions 9, 11 and 13, which form a solvent-exposed cluster in wild-type Arc (Figure 1B), mediate formation of a hydrophobic oligomer interface in a β -sheet rich structure. Both Arc-N11L and S-VLV have higher thermal folding stability than the wild type despite lower conformational specificity^{15,20}.

The ability of hydrophobic substitutions to populate additional folded structures almost certainly depends on the specific properties and sequence position of the substituted amino acids, and not solely on the number of substitutions. This is known to be true of fold switching in Arc dimers. Hydrophobic substitutions at position 11 other than leucine likewise permit formation of the alternate Arc dimer fold by enabling burial of that position; only the leucine substitution, however, leads to an alternate fold with a stability that is comparable with the wild-type conformation²¹. Arc-N11I adopts only the wild-type dimer structure despite the comparable hydrophobicity of leucine and isoleucine, possibly because

leucine has a higher α -helical propensity. Single substitutions at positions 9 and 13 have not been tested, but are not expected to have significant effect on dimer structure, since these positions do not change solvent exposure significantly in the alternate fold.

The specific sequence determinants of oligomer formation in S-VLV are considerably more mysterious, because the structure of the octamer is unknown. Oligomerization could simply be a nonspecific consequence of increased hydrophobicity, or it could depend on specific changes that may or may not require all three substitutions. To address these possibilities, we present a comprehensive study of the mutational effects of multiple hydrophobic substitutions on the surface of the Arc repressor β -sheet. Our findings show that substitution of all three positions with highly hydrophobic residues is necessary and sufficient for oligomerization, and also that different hydrophobic residues can lead to different oligomeric states. While increased hydrophobicity tends to increase structural degeneracy, the precise way in which increased degeneracy manifests itself depends on specific amino-acid properties such as side-chain branching.

MATERIALS AND METHODS

Mutagenesis, protein expression and affinity purification

A plasmid encoding the Arc S-VLV variant was previously constructed by mutagenesis of the synthetic *arc-st11* gene in plasmid pET800^{20,22}. The *arc-st11* gene includes the coding sequence for the st11 C-terminal extension (H₆KNQHE) to stabilize the proteins against degradation during expression and to allow for affinity purification²³. All other variants were constructed by QuikChange (Stratagene; Santa Clara, CA) site-directed mutagenesis of either this plasmid or the original wild-type version of pET800. The Arc repressor variants were overexpressed in LB medium in *Escherichia coli* strain BL21(λ DE3) cells transformed with the appropriate pET800 vector, and purified by denaturing Ni²⁺-NTA affinity chromatography by loading cleared cell lysates in 0.1 M NaH₂PO₄ (pH 8.0), 0.01 M Tris, 6 M guanidine hydrochloride, 0.01 M imidazole onto 3 mL Ni²⁺-NTA resin per liter of culture, washing with a total of 100 mL of the same buffer, and eluting with 6 M guanidine hydrochloride containing 0.2 M acetic acid. Eluates containing purified protein were then dialyzed into SB250 buffer [0.05 M Tris (pH 7.5), 0.25 M KCl, 0.2 mM EDTA] for 48 h at 4 °C, with one change of dialysis jar after 24 h.

Size exclusion chromatography

For analysis of oligomer formation, affinity-purified, refolded Arc repressor variants were loaded onto a Superdex 75 HR 10/300 size exclusion column (GE Healthcare), at 300 μ M protein concentration and 0.5 mL injection volume unless otherwise indicated, and eluted isocratically with SB250 buffer. Heat annealed samples were prepared by heating to 80 °C for 2 to 4 h (as described in the text) in a hot water bath, cooling to room temperature, and centrifuging at 13000 *g* for 10 min to remove any gross precipitates formed during heating. Samples thus prepared were then immediately injected onto the size exclusion column. The size exclusion column was calibrated with a set of four standards of known molecular weight (Figure S1).

Circular dichroism

Far ultraviolet circular dichroism spectra of Arc repressor variants were obtained in SB250 buffer at 20 °C in a 1.0 mm pathlength cuvette on a Olis DSM-20 CD spectrometer, using 25-50 μ M Ni-affinity purified protein. Proteins were not purified by size exclusion to allow multiple oligomeric states to be present in spectra. Heated triple substitution variants were prepared as described above, and analyzed by circular dichroism within 2 h of heat treatment and cooling. Scans were collected using a 30 s integration time per scan and signal averaging of three scans measured at 1 nm increments from 260 to 205 nm. Thermal denaturation curves were obtained in SB250 buffer using 25-50 μ M Ni-affinity purified protein (single, double, and triple substitutions in the β -strand) in a 2.0 mm pathlength cuvette. Samples were heated and cooled between 20–100 °C with a 2 °C step size and 2 min sample equilibration time at each temperature. At each temperature, ellipticity was monitored at 222 nm using 60 s total data collection time, and 50 s integration time. Thermal melt fitting was performed as described²⁴ assuming a two-state model and a C_p value of 896 cal/mol*K was set for all variants (14 cal/mol*K per residue)²⁵.

NMR spectroscopy

Uniformly 15 N-labelled Arc repressor double substitution variants were expressed as described above, but using M9T medium with 0.8 g/L 15 N-labelled ammonium chloride as sole nitrogen source. Labelled variants were purified under denaturing conditions as described above, refolded by dialysis into 0.05 M MES (pH 5.5), 0.05 M KCl, and further purified by size exclusion chromatography in the same buffer with a Superdex 75 HR 10/300 column (GE Healthcare). NMR samples were then prepared by addition of 10% $^2\text{H}_2\text{O}$ (v/v) and 0.01% sodium azide (w/v). Unless otherwise indicated, all spectra were obtained at 25 °C on a Varian Inova 600 MHz spectrometer with a cryogenic probe. Sequence-specific resonance assignments were obtained from analysis of 3D ^{15}N - ^1H NOESY-HSQC and 3D ^{15}N - ^1H TOCSY-HSQC spectra. NMR data were processed with NMRPipe²⁶ and analyzed with SPARKY (UCSF).

Analysis of sequence hydrophobicity and aggregation propensity

Sequences of Arc repressor variants were submitted to the TANGO server (<http://tango.crg.es/> L. Serrano, EMBL). Aggregation propensity is predicted based on a variety of factors including β -strand propensity (higher likelihood to aggregate), overall charge (lower charge is more aggregation-prone), and residue burial in a two-window sampling approximation²⁷. Average aggregation propensity values in Table 1 represent an average of the individual aggregation propensities for residues 8–14 given in the TANGO output file. Sequence hydrophobicity was computed as grand average of hydropathicity (GRAVY) using the program ProtParam on the ExPASy server (<http://web.expasy.org/protparam/>). GRAVY is based on Kyte-Doolittle hydropathy values²⁸.

Native mass spectrometry

Arc variant preparations were diluted to 50 μ M concentration (total protein basis) before buffer exchange into 100–300 mM ammonium acetate (pH 7.2) using Micro Bio-Spin 6 Chromatography columns (Bio-Rad Laboratories, CA, USA). Native-like solutions were

then directly infused for MS analysis *via* nanoflow electrospray ionization (nESI) conducted in static mode. All measurements were made using a Waters Synapt G2 HDMS quadrupole time of flight hybrid mass spectrometer (Manchester, UK). A nESI potential of 1.5–2.0 kV (relative to the mass spectrometer inlet) was applied directly to the solution by means of a platinum wire. Instrument parameters were tuned to ensure the preservation of noncovalent protein complexes, and to maximize the transmission and detection of ions with high m/z values. Ion source temperature was kept at approximately 25 °C (*i.e.*, ambient temperature), ion source potentials were set to their lowest practical values for each analyte, DC offsets between ion optic elements were minimized, and pressures were increased in the ion source and ion guides. MS/MS experiments were conducted using collision-induced dissociation (CID) with argon as the collision gas.

RESULTS

Sequence determinants of mutationally induced oligomerization in Arc repressor.

We expressed and purified six Arc variants containing one or two of the three hydrophobic substitutions present in S-VLV, as well as six variants containing triple substitutions of six different kinds of hydrophobic residues, including Ala, Met, Phe, Leu, Val and Ile. In this paper, we refer to these variants by the same naming convention used for S-VLV and related Arc surface variants: S indicates the β -strand region of Arc repressor (residues 9–13) and the three other letters denote the amino-acid residues found at the three surface positions of the β -strand, namely residues 9, 11 and 13^{15,20,29}.

The single and double substitutions do not convert Arc to a higher order oligomer, but instead yield dimers with increased stability against unfolding (Figure 2 and Table 1). Size exclusion traces show elution volumes similar to those of the wild-type Arc dimer. A heat-annealing treatment (at 80 °C) that quantitatively converts S-VLV to the oligomer does not convert the single and double substitutions (Figure 2A). Far ultraviolet circular dichroism spectra of the variants also do not show the loss of helicity observed for the S-VLV oligomer (Figure 2B). All variants showed sigmoidal, mostly reversible thermal denaturation curves, as monitored by circular dichroism at 222 nm (Figure S2). All single hydrophobic substitutions on the β -strand surface slightly increase the T_m of Arc (Table 1). The double hydrophobic substitutions also stabilize the Arc dimer more than the single substitutions, on average.

Even three polar-to-hydrophobic substitutions are not enough to populate oligomers unless highly nonpolar residue types are substituted. We obtained size exclusion traces for each of the six triple polar-to-hydrophobic substitutions after refolding from guanidine, and then again after 2 or 4 h of heat treatment (Figure 3). Of the six variants, the one most prone to self-association is S-III, which oligomerizes almost quantitatively even prior to heat treatment. S-VVV, S-LLL and S-FFF also oligomerize, but generally require heat treatment. These also show a reduction in helical secondary structure after heating (Figure S3), consistent with a less helical oligomer fold as seen in S-VLV²⁰. S-MMM and S-AAA do not show strong evidence for oligomerization even after extensive heating, though the dimer peak for S-MMM develops a shoulder, indicating a small quantity of higher-order species. The apparent tendency to oligomerize closely follows the Kyte-Doolittle hydrophathy index

of the amino acid: Ile (4.5) > Val (4.2) > Leu (3.8) > Phe (2.8) > Met (1.9) and Ala (1.8). The S-MMM and S-AAA dimers also have high thermal stability against unfolding ($T_m \sim 70^\circ\text{C}$) (Figure S4 and Table 1), despite forming mainly dimer as judged by size exclusion, further confirming that nonpolar substitutions generally increase the folding stability of the dimer. Thermal denaturation curves of S-VVV, S-LLL and S-FFF are broadened relative to S-MMM and S-AAA and/or show steeper lower baseline slopes (Figure S4); this presumably reflects non-two state behavior in variants that oligomerize.

Variants with different hydrophobic residues may also form Arc oligomers of differing size and/or shape. No elution volumes shorter than the S-VLV octamer or longer than the wild-type dimer are observed, although some variants show peaks with intermediate elution volumes (Figure 3). For example, the elution profiles of S-FFF contain a peak with an elution volume of ~ 12.5 mL and a heavy tail, which could theoretically result from reversible exchange between a dimer and something larger, perhaps a tetramer³⁰. The elution profiles of S-LLL contain a peak with an elution volume of 10–11 mL (corresponding approximately to hexamer) and a broad shoulder at shorter elution volume, suggesting the presence of multiple species larger than a dimer. Native mass spectrometry (MS) data for S-LLL and S-FFF variants reveal detectable tetramers in unheated samples and increased amounts of tetramer and other higher order species in heated samples (Figures S5, S6). In the case of S-LLL, the stoichiometry of tetrameric oligomer was verified by MS measurement (Figure S5; charge deconvoluted MW of 30516.59 $\pm$ 1.080 Da or 4.000 $\pm$ 0.0 subunits) and by MS/MS measurement in which the tetramer was dissociated into a trimer and released monomer (Figure S7). In the case of S-FFF both hexamer (charge deconvoluted MW of 46466.0 $\pm$ 0.0 Da or 6.011 $\pm$ 0.0 subunits) and a putative decamer (charge deconvoluted 77310.60 $\pm$ 13.97 Da or 10.001 $\pm$ 0.0 subunits) were observed (Figure S6), where the hexamer was found to undergo MS/MS decomposition to a pentamer and released monomer (Figure S7). We conclude that some of the triple hydrophobic variants form oligomers other than the octamer previously observed for S-VLV. It is noteworthy that S-VLV itself can form heptamers as well as octamers²⁰.

We also tested the effect of substitutions at Pro 8 on oligomerization of S-VLV (Figure 4). The substitutions P8L and P8A are known to strongly increase the free energy of unfolding of the wild-type Arc dimer (+2.5–2.9 kcal/mol) by expanding the hydrogen bond network of the β -sheet^{22,31}; both P8L and P8A, however, also increase the hydrophobicity of the 8–14 region (Table 1), which might be expected to favor oligomer formation based on the results presented above. VLV-P8A does not undergo oligomerization even after prolonged heat treatment (Figure 4A), and unfolds (reversibly) only at very high temperature ($T_m > 90^\circ\text{C}$; Figure 4B), suggesting that the P8A substitution stabilizes the dimer relative to the oligomer. VLV-P8L, by contrast, does oligomerize upon heat annealing (Figure 4A), despite the fact that P8L stabilizes the wild-type Arc dimer by approximately the same amount as P8A^{22,31}. VLV-P8L shows a biphasic thermal unfolding transition (Figure 4B): the first phase, which is irreversible, may represent conversion of dimer to oligomer with loss of helicity; the second phase likely represents reversible global unfolding of oligomer(s) with $T_m \sim 90^\circ\text{C}$, similar to S-VLV²⁰. We suggest that the P8L variant can oligomerize because a large aliphatic side chain at residue 8 makes stabilizing interactions in the oligomer, which offsets the gain of backbone hydrogen bonding in the dimer.

We conclude that population of higher order oligomers in Arc requires three substitutions of highly hydrophobic residue types on the surface of the β -sheet, presumably because these residues form a hydrophobic interface in the oligomer (see Figure 1). Formation of oligomers is also modulated somewhat by interactions involving residue 8 at the N terminus of the wild-type β -strand region. When the threshold of three surface substitutions is reached, oligomer is formed despite the fact that dimer folding stability also increases. These surface substitutions simultaneously increase native state stability and lower structural specificity.

Failure of hydrophobic variants to aggregate.

None of the variants studied above undergoes gross aggregation, even though every variant that undergoes oligomerization also contains a hydrophobic stretch of sequence that is predicted by TANGO²⁷ to act as a hotspot for β -aggregation (Table 1). Oligomer formation in the Arc variants may be promoted by high β -aggregation propensity, but the self-association is limited to the formation of relatively small oligomers. It is possible that these variants form closed structures such as rings that prevent formation of higher-order aggregates.

It is also possible that residues Pro 8 and Pro 15, which flank the β -strand region, act as “aggregation gatekeeper” residues that hinder aggregation even when the intervening sequence is highly aggregation-prone³². Proline and charged amino acids are selectively conserved at the flanks of aggregation-prone regions in proteins despite reducing stability, suggesting that they function to prevent aggregation³³. Proline has a relatively rigid backbone structure that disfavors β -strand conformation.

We attempted to test this idea in Arc by examining a P8A/P15A double substitution. In a wild-type Arc background, the P8A/P15A substitutions have little effect on expression level, structure, thermal stability or folding reversibility (Figure 5). An S-VLV P8A/P15A variant, however, showed severe arrest of cell growth after induction, expressed poorly, and was isolated only in very low yield after nickel affinity purification and refolding by dialysis. This protein had a far ultraviolet CD spectrum resembling an Arc dimer following refolding by dialysis, but precipitated catastrophically upon heating to 50 °C (Figure 5). These observations provide some degree of support for the hypothesis that the two prolines together act as gatekeepers against aggregation.

Effect of hydrophobic substitutions on dimer structure.

In addition to differently folded oligomers, hydrophobic substitutions may lead to changes in Arc dimer structure, such as conversion between the wild-type and the more helical dimer forms, which differ by replacement of β -sheet with helix¹⁹. An N11L substitution (S-QLR) is known to be sufficient for the more helical dimer to predominate by a roughly 3:1 ratio at 25 °C^{15,20}. The S-VLV triple substitution also retains the ability to form the helical Arc dimer, though in this case it is disfavored relative to the wild-type β -sheet by an approximately 1:3 ratio under the conditions tested²⁰.

In the present study, the two double hydrophobic substitution variants with Leu at position 11, S-QLV and S-VLR, show minor increases in helical CD signal relative to wild-type Arc,

similar to what is observed in S-QLR (see Figure 2B). Variants lacking Leu 11, including S-QNV, S-VNR and S-VNV, show slightly lower helical signals. The above observations suggest that a Leu substitution at position 11 is necessary and sufficient to populate the switch Arc dimer, while Val substitutions at positions 9 and 13 have relatively little effect on the equilibrium between the wild-type Arc or helical dimer forms.

Comparisons of NMR spectra for the double substitution variants (Figure 6) support these conclusions. Certain ^{15}N and ^1H resonances are known to be sensitive indicators of the structural transition in the dimer^{15,20,21}. The Trp 14 $\epsilon 1$ resonance of S-QLR, for example, is broadened severely near 25 °C due to exchange between the wild-type and helical dimer structures, in which the ^1H chemical shift moves by ~1 ppm^{15,20}. The two double substitution variants that contain the N11L substitution, S-QLV and S-VLR, lack any visible Trp 14 $\epsilon 1$ resonance in ^{15}N - ^1H HSQC spectra at 25 °C (Figure 6A); like S-QLR, they most likely exchange between the two dimer structures. Analysis of the HN resonances of Ala 26, Val 33, and Met 42, which are also sensitive to the fold switch but undergo less broadening than Trp 14 $\epsilon 1$, suggests that in S-QLV the wild-type and helical dimer folds are about equally populated, while in S-VLR the wild-type Arc fold predominates by about a 3:1 ratio (Figure 6B). The spectrum of double substitution variant S-VNV, meanwhile, closely resembles that of wild-type Arc, most tellingly in the Trp 14 $\epsilon 1$ resonance (Figure 6A). Overall, these results indicate that valine substitutions at positions 9 and 13 cannot by themselves populate the helical dimer fold, and in the presence of Leu 11 they slightly disfavor it relative to the wild-type structure.

DISCUSSION

Increased hydrophobicity in the β -sheet of Arc repressor progressively decreases the conformational specificity of the folded state. An alternative folded dimer is populated in equilibrium with the wild-type dimer as a specific consequence of a leucine substitution at position 11²¹, and is not promoted by other hydrophobic substitutions at surface positions in the β -sheet. One or more folded higher-order oligomers (e.g. octamer of S-VLV) are then also populated in equilibrium with dimers when (and only when) both positions 9 and 13 are substituted with highly hydrophobic residues in addition to position 11. In this case, a variety of highly hydrophobic residues (e.g. Phe, Ile, Val, Leu) can promote oligomerization, and the identity of the substituted residues has some influence on apparent oligomer size. The hydrophobic substitutions also generally increase thermal stability even as they promote alternate dimeric and oligomeric folds, illustrating that the determinants of specificity are different from those of stability.

The most obvious way for increased hydrophobicity to decrease conformational specificity is by generating alternate or additional possibilities for forming a hydrophobic core. In the case of Arc-N11L, such a mechanism is directly evident from a comparison of structures of the wild-type dimer and the fold-switched form¹⁹, in which the introduced leucine switches from a solvent-exposed position in the wild-type structure to a buried position in the alternate dimer. In the case of the triple hydrophobic substitutions, the structures of the oligomers they populate are not known, though their fold is clearly different from either dimer, as a large loss of helicity is observed²⁰. Since all three substitutions are required for

oligomers to form, one imagines that the entire solvent-exposed face of Arc repressor's β -sheet region becomes buried in the oligomeric interface (see Figure 1B). Different packing arrangements in such an interface may explain why different hydrophobic residues can generate different sized oligomers.

We doubt that higher order Arc oligomers are formed by many, if any, naturally occurring members of the Arc repressor family. Sequence similarity searches did not reveal any Arc homologs with three hydrophobic substitutions at positions 9, 11 and 13 (Figure S7). Multiple Arc homologs do have fairly hydrophobic residues at both positions 11 and 13. Because all three positions directly contact DNA³⁴, however, it is unlikely that the entire surface could be rendered hydrophobic without severely damaging sequence-specific DNA recognition. In P22 Arc repressor itself, alanine mutations at any of the three positions greatly reduces DNA binding³⁵. The failure to observe triple hydrophobic substitutions in natural Arc homologs probably owes as much or more to evolutionary constraints on binding function as to avoidance of oligomerization. Indeed, if triple hydrophobic substitutions were observed in Arc homologs, it might suggest that they were evolving novel functions through oligomerization.

Soluble proteins have a roughly constant overall fraction of hydrophobic residues³⁶, and this likely reflects a delicate balance in which native state stability is maintained without compromising specificity or solubility³⁷. We have suggested previously that heightened sequence hydrophobicity may enable a continuum of alternate folding patterns, ranging from subtle native state changes to nonnative aggregates²⁰. The extreme outcome of nonnative aggregation may be particularly favored by large increases in hydrophobicity within localized regions, generating hydrophobic blocks of sequence³⁸. Evolution appears to select against hydrophobic blocks^{39,40}, and this is commonly attributed to aggregation avoidance. While the triple hydrophobic substitution Arc variants studied here contain 6-residue hydrophobic blocks with high predicted aggregation propensity, however, these sequences oligomerize instead of aggregating. Other sequence factors, such as flanking proline aggregation "gatekeepers"³² may prevent these blocks from nucleating open-ended aggregation, but native state specificity is still lost. Selection against hydrophobic blocks may reflect not just avoidance of nonnative aggregation but also avoidance of other kinds of lowered specificity for the native state, any of which might lead to loss of function or gain of toxic function.

While structural specificity is of high value in proteins, in some cases the normal function of a protein actually employs or even requires the ability to adopt more than one folding pattern^{41,42}. In fact, the ability to switch between folds may be widespread in the protein universe⁴³. Thus, the determinants of structural specificity, including sequence hydrophobicity, are relevant not only to avoidance of undesirable structural properties but also to acquisition of useful ones.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements.

This work was supported by NIH R01 grants GM066806 and GM104040 (to M.H.J.C.).

REFERENCES

1. Lattman EE, Rose GD. Protein folding--what's the question? *Proc Natl Acad Sci U S A* 1993;90(2): 439–441. [PubMed: 8421673]
2. Bolon DN, Grant RA, Baker TA, Sauer RT. Specificity versus stability in computational protein design. *Proc Natl Acad Sci U S A* 2005;102(36):12724–12729. [PubMed: 16129838]
3. Hoffmann J, Wrabl JO, Hilser VJ. The role of negative selection in protein evolution revealed through the energetics of the native state ensemble. *Proteins*. 2016;84(4):435–447. [PubMed: 26800099]
4. Dill KA. Dominant Forces in Protein Folding. *Biochemistry-Us* 1990;29(31):7133–7155.
5. Kauzmann W Some Factors in the Interpretation of Protein Denaturation. *Adv Protein Chem* 1959;14:1–63. [PubMed: 14404936]
6. Pace CN, Fu HL, Fryar KL, et al. Contribution of Hydrophobic Interactions to Protein Stability. *J Mol Biol* 2011;408(3):514–528. [PubMed: 21377472]
7. Uversky VN, Gillespie JR, Fink AL. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins*. 2000;41(3):415–427. [PubMed: 11025552]
8. Dill KA, Bromberg S, Yue KZ, et al. Principles of Protein-Folding - a Perspective from Simple Exact Models. *Protein Sci* 1995;4(4):561–602. [PubMed: 7613459]
9. Chan HS, Dill KA. Sequence Space Soup of Proteins and Copolymers. *J Chem Phys* 1991;95(5): 3775–3787.
10. Yue K, Dill KA. Lattice Models in the Study of Protein Design. *Faseb J* 1992;6(1):A265–A265.
11. Bolon DN, Mayo SL. Polar residues in the protein core of Escherichia coli thioredoxin are important for fold specificity. *Biochemistry-Us* 2001;40(34):10047–10053.
12. Hill RB, DeGrado WF. A polar, solvent-exposed residue can be essential for native protein structure. *Struct Fold Des* 2000;8(5):471–479.
13. Lumb KJ, Kim PS. A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry-Us* 1995;34(27):8642–8648.
14. Padlan EA, Love WE. Refined Crystal-Structure of Deoxyhemoglobin-S .1. Restrained Least-Squares Refinement at 3.0-Å Resolution. *J Biol Chem* 1985;260(14):8272–8279. [PubMed: 4008491]
15. Cordes MHJ, Burton RE, Walsh NP, McKnight CJ, Sauer RT. An evolutionary bridge to a new protein fold. *Nat Struct Biol* 2000;7(12):1129–1132. [PubMed: 11101895]
16. Calamai M, Taddei N, Stefani M, Ramponi G, Chiti F. Relative influence of hydrophobicity and net charge in the aggregation of two homologous proteins. *Biochemistry-Us* 2003;42(51):15078–15083.
17. Dubay KF, Pawar AP, Chiti F, Zurdo J, Dobson CM, Vendruscolo M. Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J Mol Biol* 2004;341(5):1317–1326. [PubMed: 15302561]
18. Kim W, Hecht MH. Generic hydrophobic residues are sufficient to promote aggregation of the Alzheimer's Aβ42 peptide. *Proc Natl Acad Sci U S A* 2006;103(43):15824–15829. [PubMed: 17038501]
19. Cordes MHJ, Walsh NP, McKnight CJ, Sauer RT. Solution structure of switch arc, a mutant with 3(10) helices replacing a wild-type beta-ribbon. *J Mol Biol* 2003;326(3):899–909. [PubMed: 12581649]
20. Stewart KL, Dodds ED, Wysocki VH, Cordes MH. A polymetamorphic protein. *Protein Sci* 2013;22(5):641–649. [PubMed: 23471712]
21. Anderson TA, Cordes MH, Sauer RT. Sequence determinants of a conformational switch in a protein structure. *Proc Natl Acad Sci U S A* 2005;102(51):18344–18349. [PubMed: 16344489]

22. Milla ME, Brown BM, Sauer RT. Protein stability effects of a complete set of alanine substitutions in Arc repressor. *Nat Struct Biol* 1994;1(8):518–523. [PubMed: 7664079]
23. Milla ME, Brown BM, Sauer RT. P22 Arc repressor: enhanced expression of unstable mutants by addition of polar C-terminal sequences. *Protein Sci* 1993;2(12):2198–2205. [PubMed: 8298465]
24. Becktel WJ, Schellman JA. Protein stability curves. *Biopolymers*. 1987;26(11):1859–1877. [PubMed: 3689874]
25. Myers JK, Pace CN, Scholtz JM. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci* 1995;4(10):2138–2148. [PubMed: 8535251]
26. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 1995;6(3):277–293. [PubMed: 8520220]
27. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* 2004;22(10):1302–1306. [PubMed: 15361882]
28. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 1982;157(1):105–132. [PubMed: 7108955]
29. Cordes MHJ, Sauer RT. Tolerance of a protein to multiple polar-to-hydrophobic surface substitutions. *Protein Sci* 1999;8(2):318–325. [PubMed: 10048325]
30. Yu CM, Mun S, Wang NH. Theoretical analysis of the effects of reversible dimerization in size exclusion chromatography. *J Chromatogr A* 2006;1132(1–2):99–108. [PubMed: 16887127]
31. Schildbach JF, Milla ME, Jeffrey PD, Raumann BE, Sauer RT. Crystal structure, folding, and operator binding of the hyperstable Arc repressor mutant PL8. *Biochemistry-US* 1995;34(4):1405–1412.
32. Rousseau F, Serrano L, Schymkowitz JWH. How evolutionary pressure against protein aggregation shaped chaperone specificity. *J Mol Biol* 2006;355(5):1037–1047. [PubMed: 16359707]
33. De Baets G, Van Durme J, Rousseau F, Schymkowitz J. A genome-wide sequence-structure analysis suggests aggregation gatekeepers constitute an evolutionary constrained functional class. *J Mol Biol* 2014;426(12):2405–2412. [PubMed: 24735868]
34. Raumann BE, Rould MA, Pabo CO, Sauer RT. DNA recognition by beta-sheets in the Arc repressor-operator crystal structure. *Nature*. 1994;367(6465):754–757. [PubMed: 8107872]
35. Brown BM, Milla ME, Smith TL, Sauer RT. Scanning mutagenesis of the Arc repressor as a functional probe of operator recognition. *Nat Struct Biol* 1994;1(3):164–168. [PubMed: 7656034]
36. Irback A, Sandelin E. On hydrophobicity correlations in protein chains. *Biophys J* 2000;79(5):2252–2258. [PubMed: 11053106]
37. Sandelin E On hydrophobicity and conformational specificity in proteins. *Biophys J* 2004;86(1):23–30. [PubMed: 14695246]
38. Istrail S, Schwartz R, King J. Lattice simulations of aggregation funnels for protein folding. *J Comput Biol* 1999;6(2):143–162. [PubMed: 10421520]
39. Patki AU, Hausrath AC, Cordes MH. High polar content of long buried blocks of sequence in protein domains suggests selection against amyloidogenic non-polar sequences. *J Mol Biol* 2006;362(4):800–809. [PubMed: 16935301]
40. Schwartz R, King J. Frequencies of hydrophobic and hydrophilic runs and alternations in proteins of known structure. *Protein Sci* 2006;15(1):102–112. [PubMed: 16373477]
41. Bryan PN, Orban J. Proteins that switch folds. *Curr Opin Struct Biol* 2010;20(4):482–488. [PubMed: 20591649]
42. Dishman AF, Volkman BF. Unfolding the Mysteries of Protein Metamorphosis. *ACS Chem Biol* 2018;13(6):1438–1446. [PubMed: 29787234]
43. Porter LL, Looger LL. Extant fold-switching proteins are widespread. *Proc Natl Acad Sci U S A* 2018;115(23):5968–5973. [PubMed: 29784778]

Figure 1: Polar-to-hydrophobic substitutions in the β -strand region of Arc repressor decrease folding specificity.

A) Wild-type Arc (PDB ID 1ARR) and “switch” Arc (PDB ID 1NLA) homodimers, illustrating their distinct folding patterns. Residues 9–13 are β -sheet in wild-type Arc and helix in the alternate “switch” fold. B) Residues 9, 11 and 13 form the solvent-exposed surface of the sheet. C) Ball and stick diagrams schematically depicting structure and hydrophobicity for residues 9–14. Wild-type Arc adopts a homodimeric β -sheet structure in which hydrophobic residues (black balls) face the interior of the protein and polar residues (white balls) face solvent. Arc-N11L exists in equilibrium between the wild-type fold and the “switch” dimer fold in which the β -sheet is replaced by helices. Arc-N11L lowers conformational specificity by allowing multiple viable hydrophobic core arrangements: the introduced leucine is on the solvent-exposed surface of the wild-type fold but is buried in the hydrophobic core of the “switch” dimer. Two additional substitutions (Q9V and R13V) give the S-VLV variant and render the entire sequence of this region hydrophobic. S-VLV adopts both dimer folds, and also forms an octameric species. The structure adopted by residues 9–14 in the octamer is unknown (as indicated by dashed lines) but the global fold of the octamer is considerably less helical, and therefore different from either dimer fold. S-VLV therefore has at least three distinct folds. Our working model is that residues 9, 11 and 13 form a hydrophobic interface in the octamer.

Figure 2: One or two polar-to-hydrophobic substitutions in the β -strand are insufficient to convert Arc repressor dimer to a higher-order oligomer.

(A) Superdex 75 size exclusion chromatograms of Arc single and double polar-to-hydrophobic substitutions in SB250 buffer at 300 μ M protein concentration and ambient temperature, following 4 h of a heat annealing treatment at 80 $^{\circ}$ C. See the text for an explanation of shorthand nomenclature for variants. Hydrophobic substitutions are indicated in bold. See also Figure S1 for column calibration. (B) Far ultraviolet circular dichroism spectra of the same variants, without a heat annealing treatment, in SB250 at 25 μ M protein concentration in a 1 mm path length cuvette at 20 $^{\circ}$ C.

Figure 3: Three highly hydrophobic substitutions in the β -strand are necessary for population of the Arc repressor oligomer.

Each panel shows a set of Superdex 75 size exclusion chromatograms of Arc triple polar-to-hydrophobic substitutions in SB250 buffer at 300 μ M protein concentration and ambient temperature, either without heat annealing treatment (green), or following 2 h (yellow) or 4 h (red) of annealing at 80 $^{\circ}$ C. For comparison, dashed lines show chromatograms for heated Arc-S-VLV oligomer (orange) and wild-type Arc dimer (blue). See also Figure S1 for column calibration.

Figure 4: Effect of dimer-stabilizing P8A/P8L substitutions on the oligomerization of S-VLV.

(A) Superdex 75 size exclusion chromatograms of S-VLV-P8A (purple) and S-VLV-P8L (red), in SB250 buffer at 150 μ M protein concentration, before (dashed lines) and after (solid lines) a heat annealing treatment for 4 h at 80 $^{\circ}$ C. Chromatograms of heat annealed S-VLV octamer and wild-type Arc dimer in SB250 buffer are shown for comparison (brown). See also Figure S1 for column calibration. (B) Thermal denaturation curves for S-VLV P8A and S-VLV P8L in SB250 at 25 μ M protein concentration in a 2 mm path length cuvette, monitored by circular dichroism at 222 nm.

Figure 5: Substitution of both P8 and P15 causes some higher-order aggregation in Arc-S-VLV, but does not affect dimer formation in a wild-type background.

(A) Analytical size exclusion of unheated P8A/P15A (blue) at 380 μ M in SB250 buffer (blue). Traces for Arc-S-VLV and wild-type Arc (brown) are shown for comparison. See also Figure S1 for column calibration. (B) Far ultraviolet circular dichroism spectra at 50 μ M protein concentration in SB250 buffer, in a 1 mm path length cuvette at 20 $^{\circ}$ C. (C) Thermal denaturation of P8A/P15A at 50 μ M concentration in SB250 buffer, collected from 20–100 $^{\circ}$ C with a 2 $^{\circ}$ C step size in a 2 mm path length cuvette. (D) Thermal denaturation of Arc-S-VLV P8A/P15A (no prior heat treatment) at 25 μ M concentration in SB250 buffer, collected from 20–100 $^{\circ}$ C with a 2 $^{\circ}$ C step size in a 2 mm path length cuvette.

Figure 6: Double hydrophobic substitution variants Arc-QLV and Arc-VLR adopt both the wild-type and helical dimer folds, while Arc-VNV resembles wild-type Arc.

(A) ^{15}N - ^1H HSQC spectrum of uniform ^{15}N -labeled variants at 87 μM protein concentration in 50 mM MES (pH 5.5), 50 mM KCl at 25 $^\circ\text{C}$, compared to spectra of wild-type Arc and “switch Arc” (N11L/L12N), a variant known to exclusively adopt the helical dimer fold. In the lower left quadrant of the spectrum, the Trp 14 E1 resonance of S-VNV resembles that of wild-type Arc, while this resonance is invisible for S-QLV and S-VLR, putatively due to exchange between very different resonance positions in the wild-type and helical dimer folds. (B) Chemical shifts for the amide groups of A26, V33 and M42 in S-QLV and S-VLR are intermediate between the values observed for wild-type and switch Arc, while values for

S-VNV are close to those of wild-type Arc. (C) Ribbon diagram of Arc repressor dimer structure with residues analyzed in panel B highlighted in color. These residues are in the helices of Arc and 10–20 Å away from the center of Arc's β -sheet, but show some sensitivity to the dimer switch.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1:

TANGO aggregation propensities and thermal denaturation data collected for Arc repressor strand mutants show increasing stability and aggregation propensity with increasing polar-to-hydrophobic substitutions. Substituted positions are underlined and in boldface. TANGO scores reflect aggregation propensity averaged between residues 8-14. GRAVY scores reflect grand average of hydropathy for the region from residues 8-14. The apparent oligomeric state determination is based on the size exclusion profile of heated samples.

Variant name	Sequence (8-15)	P to NP subs.	TANGO score	GRAVY score	T _m (°C) at 50 μM	oligomer?
QNR	PQFNLRWP	0	0	-1.05	63	no
QLR	PQFLLRWP	1	0	-0.01	65	no
VNR	PVFNL ^{R} WP	1	0	0.04	67	no
QNV	PQFNLY ^{W} P	1	0	0.19	69	no
VLR	PVFLL ^{R} WP	2	1	1.09	77	no
QLV	PQFLLY ^{W} P	2	23	1.22	74	no
VNV	PVFNL ^{Y} WP	2	18	1.29	68	no
AAA	PAFALA ^{W} P	3	2	1.36	71	no
MMM	PMFML ^{M} WP	3	7	1.40	70	no
FFF	PEFFLE ^{W} P	3	58	1.79	>70	yes
LLL	PLFLL ^{L} WP	3	66	2.21	>70	yes
VVV	PVFVLY ^{W} P	3	72	2.39	>70	yes
III	PIFIL ^{I} WP	3	72	2.51	>70	yes
VLV	PVFLL ^{Y} WP	3	71	2.33	>70	yes
VLV-P8A	AVFLL ^{Y} WP	3	n.d.	2.81	>90 [†]	no
VLV-P8L	LVFLL ^{Y} WP	3	n.d.	3.10	>70 [†]	yes

[†] at 25 μM protein concentration