Mutagenic Effects on Protein Folding and Stability

by

Thomas Anthony Anderson

B.A., Biochemistry
University of California, Berkeley, 1995

Submitted to the Department of Biology in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Biology

at the

Massachusetts Institute of Technology

February 2002

Signature of Author_____
Department of Biology
February 5, 2002

Certified by_____
Robert T. Sauer, Salvador E. Luria Professor of Biology
Thesis Supervisor

Accepted By_____
Alan D. Grossman, Professor of Biology
Co-Chair, Biology Graduate Committee

Mutagenic Effects on Protein Folding Stability and Specificity

by

Thomas Anthony Anderson

Submitted to the Department of Biology in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biology

Abstract

Knowing how sequence information dictates the formation of protein structure is critical for accurate prediction of structure, for *de novo* protein design, and for understanding protein folding and misfolding. Based on extensive studies in peptide and protein systems, the pattern of polar and nonpolar amino acids, complementary packing of side chains in the protein core, electrostatic interactions, turns and helix-capping motifs, and secondary-structure propensity have been shown to be important for folding and stability to varying degrees. The pattern of polar and nonpolar residues (binary pattern) in a sequence appears to be most critical for determining the gross three-dimensional fold, whereas other interactions are responsible for the details of protein structure and stability.

Residues 9-14 of wild-type P22 Arc repressor form a two-stranded β-sheet and have the binary pattern of an amphipathic β-sheet. Switching two residues in this region (NL11, LN12) results in the formation of two right-handed 3_{10}-helices and changes the binary pattern to that of an amphipathic helix. Arc NL11 has an ambiguous binary pattern and is in dynamic equilibrium between the sheet and helical structures. Characterization of mutants in which position 11 is replaced with different hydrophobic residues shows that binary pattern is primarily responsible for determining the structure of this region of Arc. In the context of an ambiguous binary pattern, however, the detailed chemical properties of the position-11 side chain dictate which of the two competing folds is preferred.

Although surface residues are typically unimportant for protein structure and stability, there are important exceptions to this rule. In Arc, Ser32 is solvent exposed and serves as the N_{cap} residue of helix B. Characterization of a library of position-32 variants shows that this position is critical both for determining the stability and activity of Arc. Except for cysteine, the stability of these variants correlates with the frequency at which the corresponding residues are found at N_{cap} positions in the protein database. Because the position-32 side chain is close to the DNA backbone in the repressor-operator complex, only variants with small, uncharged residues at this position show significant activity *in vivo* or *in vitro*.

Thesis Supervisor: Robert T. Sauer
Title: Salvador E. Luria Professor of Biology

Acknowledgments

I thank Bob Sauer for his time, thoughts, and advice. His contributions to my science education and scientific thinking cannot be quantitated. I am continuously impressed by his ability to quickly determine the key to solving almost any experimental problem encountered in the laboratory. Equally impressive is the freedom and thought he gives each person in his laboratory.

I thank committee members Uttam RajBhandary, David Bartel, Bruce Tidor, and Jamie McKnight for helpful suggestions regarding this work. I especially thank Jamie McKnight who was instrumental in obtaining all NMR spectra.

In addition, I thank many past and present members of the Sauer laboratory for numerous pertinent discussions, experimental advice, and good times especially Matt Cordes, Wali Karzai, Mike Nohaile, John Lo, Kate Smith, Peter Shivers, Shari Spector and Alessandro Senes. I especially thank Matt Cordes for spending so much of his time teaching me science, through both experiment and discussion. Wali Karzai has been an incredible source of biochemstry knowledge and useful reagents. Furthermore, I could not have survived without having had two great baymates along the way, Matt Cordes and Chris Hayes.

I thank my family for love and encouragement especially my parents, Marge Browning and Roger Anderson, as well as my brothers, Jason and Bryon.

Finally, I cannot thank Tavita Gill enough for her patience, love, and strength. There are no words to explain what she has meant and will always mean to me.

Table of Contents

Chapter 1

Critical Interactions for Protein Structure and Stability

A major goal of protein science has been to identify the interactions that determine the native three-dimensional structure and to quantify the importance of these interactions for protein stability. Studies of this type are a prelude to the development of algorithms that would allow the design of novel protein structures and/or accurately predict the native fold from protein sequence information alone. Based on phylogenetic comparisons and studies of mutant proteins and peptides, many factors are now understood to contribute to varying degrees to protein folding and stability. These include secondary-structure propensity, the binary pattern of polar and nonpolar residues in the amino-acid sequence, complementary side-chain packing in the protein core, efficient burial of hydrophobic surface, turns and helix-capping motifs, and electrostatics. Although information of this type has led to the formulation of general structural guidelines, there is still no comprehensive set of quantitative rules. Computational protein design and structure prediction, although improving, are still goals rather than realities.

One problem is that certain sequence features appear to be quite important in some cases and relatively unimportant in others. Conflicting studies of this type highlight the importance of sequence and structural context in establishing which interactions are most important in determining protein structure. In reviewing the literature, it is apparent that hydrophobic interactions are the dominant forces in protein folding (1-3), and that binary pattern may be the most critical aspect of a sequence that constrains or defines possible protein folds. In other words, a sequence must have the correct pattern of polar and nonpolar amino acids so that proper folding of the polypeptide chain places most hydrophobic side chains in the protein core and most polar side chins on the solvent-exposed protein surface. Within the context of the correct fold, secondary-structure

propensities, complementary core packing, turn and capping interactions, and other types of polar interactions play important roles in determining the precise fold and stability.

Secondary Structure Propensity. Numerous studies have probed the intrinsic propensities of different amino acids to form β-sheets, α-helices, and β-turns in protein and peptide systems. Some studies tabulate the occurrence of different amino acids in each type of secondary structure in the protein structure database (4-6). Other studies use model proteins or structured peptides and determine the effects of placing different amino acids into specific types of secondary structure. In general, there is good agreement between both types of studies (Table 1). Certain residues do occur more frequently in some elements of secondary structure than in others, but, other than the exclusion of proline from the central regions of helices, there are few absolute rules. Moreover, mutagenic studies reveal that changing just secondary-structure preference generally results in relatively small changes in stability. Although these findings indicate that secondary-structure propensity is unlikely to be the primary determinant of protein folding and stability, even small effects summed over an entire protein sequence could make important net contributions to stability.

Smith *et al.* (7) substituted all 20 amino acids at a central β-sheet host position and determined the stabilities of the resulting proteins. These stabilities showed a good general correlation with β-sheet frequencies in known protein structures. The same group then determined the interaction energy between side chains directly across from each other on the β-sheet surface and found that residues with high β-sheet frequencies were more likely to interact favorably with their cross-strand neighbors (8). These results suggest that context-dependent packing interactions may influence how frequently certain residues occur in natural β-sheets. Indeed in other mutagenic studies, substantial

differences in stability effects were observed depending on whether substitutions were made at an edge position of the β-sheet or at a center position (9-12), suggesting again that tertiary context may be as important as "intrinsic" secondary-structure propensity in choosing residues to occupy β-sheet positions in proteins.

Stability effects have been determined for substituting almost all amino acids at two different surface positions of α-helices in T4 lysozyme (13), in a helical peptide (14), and at a surface position in an α-helix of ribonuclease T1 (14). In general, the results of these studies showed good correlations with each other and with statistical studies of normalized residue frequencies in natural α-helices (13-16). In another study, however, the stability effects of 19 different amino acids substitutions at a central residue position of a monomeric α-helical peptide were poorly correlated with the statistical occurrence of these residues in protein α-helices (17), suggesting that tertiary context is at least as important as secondary structure preference in choosing residues for particular α-helical positions in real proteins. Indeed, the same amino acids had different effects on stability when they were placed on the hydrophobic or the hydrophilic faces of a designed α-helix (18), and position-dependent stability effects were also observed at five different positions of another monomeric α-helix (19).

Statistical studies show that certain residues occur frequently in β-turns (4-6). Moreover, correlations between turn propensity and protein stability have been reported (20,21). Replacement of a turn residue of transposon Tn10-encoded metal-tetracycline/H+ antiporter (TET) with various residues showed a correlation of protein activity with β-turn propensity (22). In some instances, however, the effects on structure and stability of residue substitutions in turns is highly context dependent (23,24).

Why would certain residues stabilize one type of secondary structure more than another? By definition, the main-chain dihedral angles (phi and psi) of residues in α-helices, β-strands, and turns are different. Hence, a given side chain might prefer an α-helix to a β-sheet because it makes more favorable steric contacts in the helix than the sheet or because solvation of the side chain and/or main chain is more favorable in the helical conformation (25). β-branched residues have low helical propensity because it is entropically unfavorable for β-branched residues to adopt a helical conformation (7,26). It is not entropically costly for β-branched residues to adopt a β-sheet conformation. β-branched residues may also prefer β-sheet conformations because they sterically occlude water molecules that would otherwise compete with the hydrogen bonds between adjacent β-strands (27). Residues that can adopt unusual backbone conformations are often found in turns, where they allow changes in backbone direction that would be extremely costly for other residues (6). Residues with high α-helical propensities are able to shield a greater amount of hydrophobic surface from solvent in a helical conformation than in an extended β-conformation (7,13).

Helix-Capping Motifs. In an α-helix, neither the first four main-chain –NH groups nor the last four main-chain –C=O groups make intra-helical hydrogen bonds. In addition, because all hydrogen bonds and peptide groups within an α-helix point in the same direction, the cumulative effect from the dipole of each peptide bond is a dipole moment at either end of the helix (28). As a consequence, helices should be stabilized if additional favorable interactions, mediated by side chains, can be made with the unsatisfied –NH and –C=O groups at the N-terminus and C-terminus, respectively (29), or with the dipole moments at helix termini. Indeed, negatively charged side chains are often stabilizing at the N-termini of α-helices, whereas positively charged side chains are stabilizing at the C-termini of α-helices (30,31). Statistical surveys of residues found at

the ends of α-helices have also led to the identification of specific types of N-terminal and C-terminal helix-capping motifs that involve hydrogen bonding to some of the unsatisfied –NH and –C=O groups within the helix (see Fig. 1 for examples) (29,32-36). Capping motifs are important for protein stability in an equilibrium sense and may serve to initiate and/or terminate elements of secondary structure during the kinetic process of folding.

Helix C-capping structures—including the "Schellman", α_L, "V-C^3", and Pro motifs—are distinguished by characteristic patterns of polar and nonpolar residues. Mutational analysis of such motifs in peptides and proteins show that they stabilize α-helices in some instances but not in others (32-34). Helix N-capping structures include the N-capping box, the "big" box, and the β-box (Fig. 1) (35). The N-capping box (Fig. 1), which has been most extensively studied, includes characteristic hydrogen bonds between side-chain and main-chain atoms as well as a hydrophobic interaction between a side chain in the helix and another side chain N-terminal to the helix (see Chapter 3 for further discussion). Mutagenic analysis of N-capping box interactions in peptides and proteins show that they can play substantial roles in determining native stability (37-46). In general, changing capping-motif positions from preferred to non-preferred residues is more destabilizing than changing surface positions within elements of secondary structure from favorable to unfavorable amino acids. Hence, capping-motif interactions appear to be more important than secondary-structure propensities in determining net stability. Nevertheless, capping motifs are not present in all α-helices and thus are not an essential element of structure or stability.

Turn Sequences. The role of turns in protein structure and stability is controversial. Turns separate elements of secondary structure and may stabilize proteins by accelerating

folding via a restriction of the local polypeptide conformation to a near native state (21,47). In some instances, however, mutation of turn sequences has little effect on protein structure or stability (48-51). In other cases, turn sequences are highly conserved among related proteins even though they are not required for function (50-52). This finding suggests some role in structure or stability. Moreover, certain turn sequences retain near native structures in short peptides or even in denatured proteins (53-57), and mutation of some turn sequences destabilizes peptide and protein structure (21,23,58-60). Turn formation by the same sequence can be context dependent (23), and thus, turn sequences may be more important in some structural environments, or for some fold topologies than for others.

Electrostatic Interactions. Most proteins fold with the vast majority of the polar side chains on the surface. Why then do side-chain mediated salt bridges and/or hydrogen bonds occur at buried positions in many protein cores? Buried polar interactions could serve to impart conformational specificity or uniqueness (61). The idea here is that alternative conformations in which these polar groups were still buried but did not form complementary polar interactions would be highly destabilized. In support of this model, placing polar groups into the core of a designed protein reduced its molten-globule character and resulted in a more native-like protein (62). Similarly, buried polar interactions in coiled-coils and helical bundles have been shown to determine whether these molecules form a single type of oligomeric structures (for example, a trimer) rather than a mixture of competing oligomeric forms (62,63).

Calculations suggest that most buried salt bridges destabilize proteins compared to isosteric hydrophobic interactions because the desolvation penalty imposed required to bury the polar groups is greater than the favorable electrostatic interactions of the side

chains within the low-dielectric protein interior (64). Indeed, replacing a buried salt bridge in Arc repressor with hydrophobic amino acids (Fig. 2) resulted in substantial stabilization of the protein with no significant change in its overall conformation (65). Replacing a buried asparagine with a leucine in the neutral protease of *Bacillus stearothermophilus* also caused an increase in thermostability (66). However, buried polar interactions are more prevalent in proteins from hyper-thermophilic organisms, suggesting that such interactions may contribute to the extreme stability of certain proteins (67). There is evidence that buried electrostatic interactions may contribute to the thermostability of cytochrome P450 (68). Clearly, whether buried electrostatic interactions are stabilizing or destabilizing is highly protein and context dependent, and indeed this is both an experimental and a theoretical finding (64).

Polar side-chains on the surface of a protein generally do not contribute significantly to protein stability (3,69,70). The high dielectric constant of water reduces the energy of electrostatic interactions on a protein's surface. However, there are instances where electrostatic interactions on the surface of a protein can have effects on protein stability. Surface side chains with the same charge that are close can destabilize the native protein because electrostatic repulsion is greater in the folded state than in the unfolded state. Salt bridges, involving side chains of opposite charge, can be stabilizing and protein structures have a significant number of ion pairs on the surface (71,72). Mutations which disrupt these salt bridges can destabilize proteins from 0.5-3 kcal/mol per salt bridge (73,74).

Complementary Side-Chain Packing in Protein Cores. Examination of known structures shows that side chains pack together to form protein cores without large voids or significant cavities. In general, substitution of large buried side chains with smaller

ones is destabilizing (75,76). For example, specific Leu→Ala or Phe→Ala substitutions in the core of T4 lysozyme destabilized the protein from 2.7 kcal/mol for single mutants to 8.3 kcal/mol for double mutants (76). Hydrophobic substitutions that maintain volume and core packing are often allowed (Fig. 3). Replacing small core residues with larger ones usually results in significant destabilization. In some instances, proteins tolerate large→small or small→large core substitutions reasonably well. In such cases, the structure of the mutant protein almost invariably shows that the protein structure relaxes, often via small main-chain and side-chain movements, to allow complementary side-chain packing by filling the cavity left by the smaller side chain or accommodating the larger side chain (77). This context dependence of core mutations suggests that some protein folds are more flexible or more able to adopt slightly different conformations than are other folds. Structural relaxation in some mutants but not others presumably also explains why reductions in stability correlate with the number of methylene groups removed in some cases but not in others (75,78).

Complementary side-chain packing is important for protein stability and also plays an important role in specifying the precise structure and oligomeric state of a protein. In a designed coiled coil, for example, mutation of isoleucine to leucine at a small number of core positions changed the oligomeric state of the protein in a highly specific fashion (79,80). In some cases, it has been possible to obtain well-folded proteins after mutating core residues to other hydrophobic side chains (81-84), whereas, in other cases, only the wild-type combination of core hydrophobic residues results in wild-type stability and activity (85). The hydrophobic cores of several proteins have also been successfully redesigned using computer algorithms to predict well-packed combinations of hydrophobic side chains in specific rotamer conformations (83,86-88).

Binary Patterning. The binary pattern of polar and non-polar residues in a polypeptide chain dictates which three-dimensional structures are potentially accessible. Some structures are ruled out because they would bury too many unsatisfied polar side chains. Other structures are eliminated because only a small number of hydrophobic side chains would be buried in the core, providing insufficient favorable free energy to offset the unfavorable entropic cost of folding (1-3). Many elements of secondary structure have one face exposed to solvent and the other face buried in the protein core, giving rise to a characteristic pattern of exposed and buried side chains that is often replicated in the pattern of polar and non-polar side chains. Indeed, the most common pentapeptide patterns of polar (P) and hydrophobic (H) residues in α-helices (PHPPH and HPPHH) are those expected for an amphipathic α-helix (89). By contrast, HPHPH and PHPHP patterns are frequently, but not always, associated with β-sheets with one face on the protein surface.

Small changes in binary pattern can have large effects on protein structure. In P22 Arc repressor, for instance, swapping adjacent polar and hydrophobic residues in the N-terminal portion of the molecule results in a dramatic structural transformation in which the two-stranded, antiparallel β-sheet of the wild-type protein is replaced by two right-handed 3_{10}-helices (90). Design experiments have proved to be powerful probes of the importance of binary pattern. In pioneering studies, Hecht and his colleagues showed that most members of a randomized library with the binary pattern expected for a designed four-helix bundle protein were soluble, α-helical, and stable, although most had molten-globule characteristics (91,92). This same group designed a library of sequences with the binary pattern of an amphiphilic β-sheet and demonstrated that many of these sequences formed self-assembling β-sheet monlayers at an air/water interface (93). A variant of the engrailed homeodomain, designed with binary pattern information alone,

was shown to be folded and hyperstable (94). A synthetic peptide, with a different sequence but the same binary pattern as rat growth-hormone releasing hormone, was recognized by a polyclonal antibody raised against the native peptide, bound to the same receptor, and had similar biological activity (95). Finally, the folds of 85% of a set of 195 native proteins were correctly predicted by an algorithm that used binary-pattern recognition (96). These studies suggest that binary pattern is largely sufficient to specify the three-dimensional folds of many proteins.

As discussed previously, core positions often tolerate mutations to other hydrophobic residues but rarely tolerate mutations to polar residues. Surface positions, by contrast, generally tolerate mutations to either polar residues or to hydrophobic residues (69,70,97,98). Because P→H substitutions change binary pattern, the latter observation implies that many different binary patterns must be compatible with particular protein folds. In fact, variants of T4 lysozyme (99), a TFIIIA-type zinc finger (100), a homeodomain protein (70), and Arc repressor (101) have been constructed that bear from 6 to 30 alanine substitutions at surface positions. In these cases, the alanine-substituted proteins folded stably and were active, despite having binary patterns dramatically different from the wild-type pattern. In such cases, the presence of hydrophobic residues at core positions may be significantly more important than the presence of polar residues at surface positions. In fact, identical protein sequences of 5-10 residues can fold into different secondary structures in different structural contexts (102-104), as long as hydrophobic residues are present at key core positions.

Why does binary pattern seem to be critical in some cases and relatively unimportant in other cases? The answer may depend on whether alternative binary patterns allow access to alternative three-dimensional folds. If small changes in binary

pattern allow new chain topologies, then some binary patterns may be ambiguous and permit competition between the native structure and alternative fold(s). A dramatic example of this occurs in Arc repressor where a single P→H mutation results in a protein that equilibrates rapidly between structures in which the N-terminus of the protein forms a two-stranded β-sheet or two right-handed 3_{10}-helices (105). Interestingly, most of the mutations in the PrP protein, which result in inherited prion diseases in humans, do not alter binary pattern (Table 2) (106,107). Pathogenic variants of PrP (PrPSc) result in neurodegenerative diseases such as kuru, Creutzfeldt-Jakob disease, Gerstmann-Straussler-Scheinker syndrome, and fatal familial insomnia. The point mutations which segregate with many of these inherited prion diseases in humans result in large changes in structure, with the normal cellular form (PrPC) having a high α-helix content and little or no β-sheet and the PrPSc form having significant β-sheet content and less α-helix (107). Prion disease may represent another case in which the binary pattern of a protein is ambiguous, allowing access to alternative structural forms whose populations depend on the detailed properties of the side chains at specific P or H positions. In chapter 2, I describe experiments with Arc repressor that show how seemingly minor changes in side-chain properties can have dramatic structural consequences in the context of an ambiguous binary pattern.

Insights from Protein Design. Binary pattern, complementary core packing, electrostatic interactions, secondary structure propensities, turn sequences, and helix-capping motifs have been shown to be important for the folding and stability of specific proteins. Which factor is the most important in determining protein structure? Which factor is least important? It would be useful to understand the relative importance of each of these factors, and several studies have attempted to address this issue. Binary pattern together with complementary side-chain packing in the protein core almost always appear

to be very important factors, but the relative importance of the other factors and interactions varies from case to case and may simply depend on the structural context of the particular protein under investigation. One obvious problem in making such comparisons is that changing an amino-acid sequence in a mutagenesis or design experiment almost always changes multiple factors at the same time.

A four-helix bundle protein, α4, designed using a simplified hydrophobic core composed only of leucines was found to be very stable (108). However, this protein lacked well-resolved side chain resonances in the NMR spectra and rates of proton-deuteron exchange were intermediate between those observed for α-helices in native proteins and in protein molten globules. Replacement of some of the hydrophobic interactions with metal-ligand interactions created protein with significantly increased native character (108). In this case, buried polar interactions appear to play a primary role in determining structural uniqueness. The design of the β-sheet protein, betabellin 14D, included a binary pattern of alternating polar and nonpolar residues plus preferred β-turn sequences but the resulting protein did not fold stably (109,110). In this case, introduction of an intra-chain disulfide bond was sufficient to permit stable folding of this designed protein (111). This example shows that binary pattern, by itself, may be sufficient to define a given fold but that this fold may not be significantly populated in the absence of other stabilizing interactions.

The relative importance of binary pattern versus secondary-structure propensity was investigated by examining the structure of four peptides (112). Two peptides were designed and constructed which had the binary pattern of an amphipathic β-sheet. One peptide was composed of residues with high helical propensity, and the other was composed of residues with high sheet propensity. Two other peptides had the binary pattern of an amphipathic α-helix. Again, one of these peptides was composed of

residues with high α-helical propensity, and the other was composed of residues with high β-sheet propensity. When the structures of each peptide were determined, those with a helical binary pattern were α–helical and those with a sheet binary pattern formed β-sheets, regardless of the secondary structure propensities of the residues used to establish the binary pattern.

Thesis Organization. In chapters 2 and 3, I present experiments, using the Arc repressor system, that probe two sequence positions that play critical roles in determining the folding specificity and stability of this protein. Arc is a 53 residue protein that folds as a homodimer with both subunits contributing to a single hydrophobic core. The structure of the native dimer (106 residues), as determined by solution NMR and x-ray crystallography (113,114), consists of a relatively globular structure formed from a two-stranded antiparallel β-sheet and four α-helices. Arc functions as a repressor of gene expression by binding to a 21-base-pair operator site in the genome of phage P22 (115). The crystal structure of the Arc-operator complex has been determined (114) and extensive mutagenesis experiments have explored the roles of individual residues in Arc in determining its activity, thermodynamic stability, and folding kinetics (65,85,90,98,105,116-123).

Prior studies from this lab had shown that Arc was remarkably tolerant to sequence substitutions on the protein surface (98,101,119). For example, an Arc mutant with 15 simultaneous Xaa→Ala mutations at surface positions (30 in the native homodimer) was shown to be more stable than wild-type and to retain biological activity. Nevertheless, two surface positions in Arc stand out as playing important roles in structure and/or stability. The first is Asn11, the central surface residue in the two-stranded antiparallel β-sheet of Arc. Cordes *et al.* (105) showed that mutation of this

position to Leu11 created an ambiguous binary pattern that resulted in a molecule that could access both the wild-type fold and the "switch" Arc fold, in which the β-sheet is replaced with two 3_{10}-helices. In chapter 2, I examine the effects of substituting other hydrophobic residues at position 11. I find that the chemical identity of the residue-11 side chain has a dramatic effect in determining whether the wild-type or "switch" fold predominates in the background of the ambiguous binary pattern. When this ambiguity is removed by addition of the Leu12→Asn substitution, however, then all hydrophobic substitutions at position 11 result in proteins that stably adopt the "switch" structure.

Although Ser32 is a surface residue in wild-type Arc, substitutions with other residues did not appear to be allowed in randomization experiments (98) and site-directed replacement of this residue with alanine was very destabilizing (119). In chapter 3, I describe the construction and characterization of a library of position-32 substitutions. These studies show that the stability of the resulting mutants correlates with their ability to function as helix N-capping residues, whereas their activity in operator-DNA binding assays correlates with side-chain size. As a result, the identity of the position-32 side chain in Arc influences both stability and biological function, and the wild-type residue, Ser32, represents a compromise between these two factors.

References

1. Kauzmann, W. (1959) *Adv. Protein Chem.* **14**(1)
2. Anfinsen, C. B. (1973) *Science* **181**(96), 223-30.
3. Dill, K. A. (1990) *Biochemistry* **29**(31), 7133-55.
4. Chou, P. Y., and Fasman, G. D. (1978) *Adv Enzymol Relat Areas Mol Biol* **47**, 45-148.
5. Deleage, G., and Roux, B. (1987) *Protein Eng* **1**(4), 289-94.
6. Levitt, M. (1978) *Biochemistry* **17**(20), 4277-85.
7. Smith, C. K., Withka, J. M., and Regan, L. (1994) *Biochemistry* **33**(18), 5510-7.
8. Smith, C. K., and Regan, L. (1995) *Science* **270**(5238), 980-2.
9. Finkelstein, A. V., and Reva, B. A. (1991) *Nature* **351**(6326), 497-9.
10. Garratt, R. C., Thornton, J. M., and Taylor, W. R. (1991) *FEBS Lett* **280**(1), 141-6.
11. Wouters, M. A., and Curmi, P. M. (1995) *Proteins* **22**(2), 119-31.
12. Minor, D. L., Jr., and Kim, P. S. (1994) *Nature* **371**(6494), 264-7.
13. Blaber, M., Zhang, X. J., and Matthews, B. W. (1993) *Science* **260**(5114), 1637-40.
14. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1997) *Proc Natl Acad Sci U S A* **94**(7), 2833-7.
15. Avbelj, F., Luo, P., and Baldwin, R. L. (2000) *Proc Natl Acad Sci U S A* **97**(20), 10786-91.
16. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1997) *Biochemistry* **36**(36), 10923-9.
17. Merutka, G., Lipton, W., Shalongo, W., Park, S. H., and Stellwagen, E. (1990) *Biochemistry* **29**(32), 7511-5.

18. Zhou, N. E., Kay, C. M., Sykes, B. D., and Hodges, R. S. (1993) *Biochemistry* **32**(24), 6190-7.

19. Petukhov, M., Munoz, V., Yumoto, N., Yoshikawa, S., and Serrano, L. (1998) *J Mol Biol* **278**(1), 279-89.

20. Jessen-Marshall, A. E., Paul, N. J., and Brooker, R. J. (1995) *J Biol Chem* **270**(27), 16251-7.

21. Mok, Y. K., Elisseeva, E. L., Davidson, A. R., and Forman-Kay, J. D. (2001) *J Mol Biol* **307**(3), 913-28.

22. Yamaguchi, A., Kimura, T., Someya, Y., and Sawai, T. (1993) *J Biol Chem* **268**(9), 6496-504.

23. Chen, P. Y., Lin, C. K., Lee, C. T., Jan, H., and Chan, S. I. (2001) *Protein Sci* **10**(9), 1794-800.

24. Aroeti, B., Kosen, P. A., Kuntz, I. D., Cohen, F. E., and Mostov, K. E. (1993) *J Cell Biol* **123**(5), 1149-60.

25. Munoz, V., and Serrano, L. (1994) *Proteins* **20**(4), 301-11.

26. Creamer, T. P., and Rose, G. D. (1992) *Proc Natl Acad Sci U S A* **89**(13), 5937-41.

27. Bai, Y., and Englander, S. W. (1994) *Proteins* **18**(3), 262-6.

28. Creighton, T. E. (1993) *Proteins: Structures and Molecular Properties*, Second Ed., W. H. Freeman and Company, New York

29. Presta, L. G., and Rose, G. D. (1988) *Science* **240**(4859), 1632-41.

30. Richardson, J. S., and Richardson, D. C. (1988) *Science* **240**(4859), 1648-52.

31. Doig, A. J., and Baldwin, R. L. (1995) *Protein Sci* **4**(7), 1325-36.

32. Viguera, A. R., and Serrano, L. (1995) *J Mol Biol* **251**(1), 150-60.

33. Prieto, J., and Serrano, L. (1997) *J Mol Biol* **274**(2), 276-88.

34. Aurora, R., Srinivasan, R., and Rose, G. D. (1994) *Science* **264**(5162), 1126-30.

35. Aurora, R., and Rose, G. D. (1998) *Protein Sci* **7**(1), 21-38.

36. Doig, A. J., MacArthur, M. W., Stapley, B. J., and Thornton, J. M. (1997) *Protein Sci* **6**(1), 147-55.

37. Lyu, P. C., Gans, P. J., and Kallenbach, N. R. (1992) *J Mol Biol* **223**(1), 343-50.

38. Lyu, P. C., Wemmer, D. E., Zhou, H. X., Pinker, R. J., and Kallenbach, N. R. (1993) *Biochemistry* **32**(2), 421-5.

39. Zhou, H. X., Lyu, P., Wemmer, D. E., and Kallenbach, N. R. (1994) *Proteins* **18**(1), 1-7.

40. Zhukovsky, E. A., Mulkerrin, M. G., and Presta, L. G. (1994) *Biochemistry* **33**(33), 9856-64.

41. Serrano, L., and Fersht, A. R. (1989) *Nature* **342**(6247), 296-9.

42. elMasry, N. F., and Fersht, A. R. (1994) *Protein Eng* **7**(6), 777-82.

43. Cocco, R., Stenberg, G., Dragani, B., Rossi Principe, D., Paludi, D., Mannervik, B., and Aceto, A. (2001) *J Biol Chem* **276**(34), 32177-83.

44. Munoz, V., Blanco, F. J., and Serrano, L. (1995) *Nat Struct Biol* **2**(5), 380-5.

45. Munoz, V., and Serrano, L. (1995) *Biochemistry* **34**(46), 15301-6.

46. Chakrabartty, A., Doig, A. J., and Baldwin, R. L. (1993) *Proc Natl Acad Sci U S A* **90**(23), 11332-6.

47. Cordes, M. H., Davidson, A. R., and Sauer, R. T. (1996) *Curr Opin Struct Biol* **6**(1), 3-10.

48. Schultz, L. W., Hargraves, S. R., Klink, T. A., and Raines, R. T. (1998) *Protein Sci* **7**(7), 1620-5.

49. Predki, P. F., and Regan, L. (1995) *Biochemistry* **34**(31), 9834-9.

50. Brunet, A. P., Huang, E. S., Huffine, M. E., Loeb, J. E., Weltman, R. J., and Hecht, M. H. (1993) *Nature* **364**(6435), 355-8.

51. Castagnoli, L., Vetriani, C., and Cesareni, G. (1994) *J Mol Biol* **237**(4), 378-87.

52. O'Neill, J. W., Kim, D. E., Johnsen, K., Baker, D., and Zhang, K. Y. (2001) *Structure (Camb)* **9**(11), 1017-27.

53. Ilyina, E., Milius, R., and Mayo, K. H. (1994) *Biochemistry* **33**(45), 13436-44.

54. Zhang, O., and Forman-Kay, J. D. (1995) *Biochemistry* **34**(20), 6784-94.

55. Blanco, F. J., Rivas, G., and Serrano, L. (1994) *Nat Struct Biol* **1**(9), 584-90.

56. Shin, H. C., Merutka, G., Waltho, J. P., Wright, P. E., and Dyson, H. J. (1993) *Biochemistry* **32**(25), 6348-55.

57. Yao, J., Feher, V. A., Espejo, B. F., Reymond, M. T., Wright, P. E., and Dyson, H. J. (1994) *J Mol Biol* **243**(4), 736-53.

58. Chen, P. Y., Gopalacushina, B. G., Yang, C. C., Chan, S. I., and Evans, P. A. (2001) *Protein Sci* **10**(10), 2063-74.

59. Sahu, A., Soulika, A. M., Morikis, D., Spruce, L., Moore, W. T., and Lambris, J. D. (2000) *J Immunol* **165**(5), 2491-9.

60. Labrou, N. E., Mello, L. V., and Clonis, Y. D. (2001) *Eur J Biochem* **268**(14), 3950-7.

61. Rozwarski, D. A., Gronenborn, A. M., Clore, G. M., Bazan, J. F., Bohm, A., Wlodawer, A., Hatada, M., and Karplus, P. A. (1994) *Structure* **2**(3), 159-73.

62. Raleigh, D. P., and DeGrado, W. F. (1992) *J Am Chem Soc* **114**, 10079-81

63. Lumb, K. J., and Kim, P. S. (1995) *Biochemistry* **34**(27), 8642-8.

64. Hendsch, Z. S., and Tidor, B. (1994) *Protein Sci* **3**(2), 211-26.

65. Waldburger, C. D., Schildbach, J. F., and Sauer, R. T. (1995) *Nat Struct Biol* **2**(2), 122-8.

66. Eijsink, V. G., van der Zee, J. R., van den Burg, B., Vriend, G., and Venema, G. (1991) *FEBS Lett* **282**(1), 13-6.

67. Burkhard, P., Stetefeld, J., and Strelkov, S. V. (2001) *Trends Cell Biol* **11**(2), 82-8.

68. Maves, S. A., and Sligar, S. G. (2001) *Protein Sci* **10**(1), 161-8.

69. Zhang, X. J., Baase, W. A., and Matthews, B. W. (1991) *Biochemistry* **30**(8), 2012-7.

70. Shang, Z., Isaac, V. E., Li, H., Patel, L., Catron, K. M., Curran, T., Montelione, G. T., and Abate, C. (1994) *Proc Natl Acad Sci U S A* **91**(18), 8373-7.

71. Wada, A., and Nakamura, H. (1981) *Nature* **293**(5835), 757-8.

72. Barlow, D. J., and Thornton, J. M. (1983) *J Mol Biol* **168**(4), 867-85.

73. Fersht, A. R. (1985) *J. Mol. Biol.* **64,** 497-

74. Perutz, M. F., and Raidt, H. (1975) *Nature* **255,** 256-

75. Shortle, D., Stites, W. E., and Meeker, A. K. (1990) *Biochemistry* **29**(35), 8033-41.

76. Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P., and Matthews, B. W. (1992) *Science* **255**(5041), 178-83.

77. Eriksson, A. E., Baase, W. A., and Matthews, B. W. (1993) *J Mol Biol* **229**(3), 747-69.

78. Kellis, J. T., Jr., Nyberg, K., and Fersht, A. R. (1989) *Biochemistry* **28**(11), 4914-22.

79. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) *Science* **262**(5138), 1401-7.

80. Harbury, P. B., Kim, P. S., and Alber, T. (1994) *Nature* **371**(6492), 80-3.

81. Lim, W. A., and Sauer, R. T. (1991) *J Mol Biol* **219**(2), 359-76.

82. Fredericks, Z. L., and Pielak, G. J. (1993) *Biochemistry* **32**(3), 929-36.

83. Desjarlais, J. R., and Handel, T. M. (1995) *Protein Sci* **4**(10), 2006-18.

84. Munson, M., O'Brien, R., Sturtevant, J. M., and Regan, L. (1994) *Protein Sci* **3**(11), 2015-22.

85. Milla, M. E., and Sauer, R. T. (1995) *Biochemistry* **34**(10), 3344-51.

86. Desjarlais, J. R., and Clarke, N. D. (1998) *Curr Opin Struct Biol* **8**(4), 471-5.

87. Dahiyat, B. I., Sarisky, C. A., and Mayo, S. L. (1997) *J Mol Biol* **273**(4), 789-96.

88. Dahiyat, B. I., and Mayo, S. L. (1997) *Science* **278**(5335), 82-7.

89. West, M. W., and Hecht, M. H. (1995) *Protein Sci* **4**(10), 2032-9.

90. Cordes, M. H., Walsh, N. P., McKnight, C. J., and Sauer, R. T. (1999) *Science* **284**(5412), 325-8.

91. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., and Hecht, M. H. (1993) *Science* **262**(5140), 1680-5.

92. Roy, S., and Hecht, M. H. (2000) *Biochemistry* **39**(16), 4603-7.

93. Xu, G., Wang, W., Groves, J. T., and Hecht, M. H. (2001) *Proc Natl Acad Sci U S A* **98**(7), 3652-7.

94. Marshall, S. A., and Mayo, S. L. (2001) *J Mol Biol* **305**(3), 619-31.

95. Weigent, D. A., Clarke, B. L., and Blalock, J. E. (1994) *Immunomethods* **5**(2), 91-7.

96. Huang, E. S., Subbiah, S., and Levitt, M. (1995) *J Mol Biol* **252**(5), 709-20.

97. Reidhaar-Olson, J. F., and Sauer, R. T. (1990) *Proteins* **7**(4), 306-16

98. Bowie, J. U., and Sauer, R. T. (1989) *Proc Natl Acad Sci U S A* **86**(7), 2152-6.

99. Heinz, D. W., Baase, W. A., and Matthews, B. W. (1992) *Proc Natl Acad Sci U S A* **89**(9), 3751-5.

100. Michael, S. F., Kilfoil, V. J., Schmidt, M. H., Amann, B. T., and Berg, J. M. (1992) *Proc Natl Acad Sci U S A* **89**(11), 4796-800.

101. Brown, B. M., Milla, M. E., Smith, T. L., and Sauer, R. T. (1994) *Nat Struct Biol* **1**(3), 164-8.

102. Cohen, B. I., Presnell, S. R., and Cohen, F. E. (1993) *Protein Sci* **2**(12), 2134-45.

103. Kabsch, W., and Sander, C. (1984) *Proc Natl Acad Sci U S A* **81**(4), 1075-8.

104. Minor, D. L., Jr., and Kim, P. S. (1996) *Nature* **380**(6576), 730-4.

105. Cordes, M. H., Burton, R. E., Walsh, N. P., McKnight, C. J., and Sauer, R. T. (2000) *Nat Struct Biol* **7**(12), 1129-32.

106. Huang, Z., Gabriel, J. M., Baldwin, M. A., Fletterick, R. J., Prusiner, S. B., and Cohen, F. E. (1994) *Proc Natl Acad Sci U S A* **91**(15), 7139-43.

107. Harrison, P. M., Bamborough, P., Daggett, V., Prusiner, S. B., and Cohen, F. E. (1997) *Curr Opin Struct Biol* **7**(1), 53-9.

108. Handel, T. M., Williams, S. A., and DeGrado, W. F. (1993) *Science* **261**(5123), 879-85.

109. Erickson, B. W., Daniels, S. B., Reddy, P. A., Unson, C. G., Richardson, J. S., and Richardson, D. C. (1986) in *Cold Spring Harbor Symposium* (Fletterick, R., and Zoller, M., eds), pp. 53-7, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

110. McClain, R. D. (1991), University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

111. Yan, Y., and Erickson, B. W. (1994) *Protein Sci* **3**(7), 1069-73.

112. Xiong, H., Buckwalter, B. L., Shieh, H. M., and Hecht, M. H. (1995) *Proc Natl Acad Sci U S A* **92**(14), 6349-53.

113. Breg, J. N., van Opheusden, J. H., Burgering, M. J., Boelens, R., and Kaptein, R. (1990) *Nature* **346**(6284), 586-9.

114. Raumann, B. E., Rould, M. A., Pabo, C. O., and Sauer, R. T. (1994) *Nature* **367**(6465), 754-7.

115. Vershon, A. K., Youderian, P., Susskind, M. M., and Sauer, R. T. (1985) *J Biol Chem* **260**(22), 12124-9.

116. Bowie, J. U., and Sauer, R. T. (1989) *Biochemistry* **28**(18), 7139-43.

117. Waldburger, C. D., Jonsson, T., and Sauer, R. T. (1996) *Proc Natl Acad Sci U S A* **93**(7), 2629-34.

118. Milla, M. E., and Sauer, R. T. (1994) *Biochemistry* **33**(5), 1125-33.

119. Milla, M. E., Brown, B. M., and Sauer, R. T. (1994) *Nat Struct Biol* **1**(8), 518-23.

120. Milla, M. E., Brown, B. M., Waldburger, C. D., and Sauer, R. T. (1995) *Biochemistry* **34**(42), 13914-9.

121. Sauer, R. T., Milla, M. E., Waldburger, C. D., Brown, B. M., and Schildbach, J. F. (1996) *Faseb J* **10**(1), 42-8.

122. Jonsson, T., Waldburger, C. D., and Sauer, R. T. (1996) *Biochemistry* **35**(15), 4795-802.

123. Cordes, M. H., and Sauer, R. T. (1999) *Protein Sci* **8**(2), 318-25.

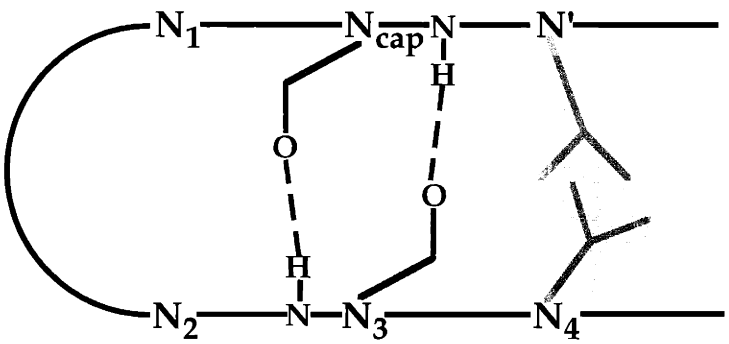
124. Nooren, I. M., Rietveld, A. W., Melacini, G., Sauer, R. T., Kaptein, R., Boelens, R. (1999) *Biochemistry* **38**(19), 6035-42.

125. Bonvin, A. M., Vis, H., Breg, J. N., Burgering, M. J., Boelens, R., Kaptein, R. (1994) *J Mol Biol* **236**(1), 328-41.

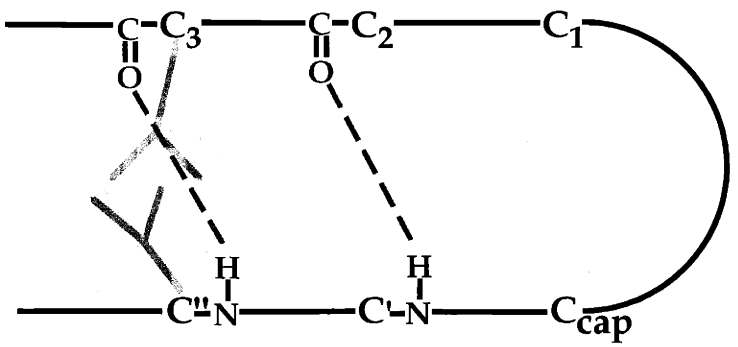
126. Heinz, D. W., Baase, W. A., Dahlquist, F. W., Matthews, B. W. (1993) *Nature* **361**(6412), 561-4.

127. Baldwin, E. P., Hajiseyedjavadi, O., Baase, W. A., Matthews, B. W. (1993) *Science* **262**(5140), 1715-8.

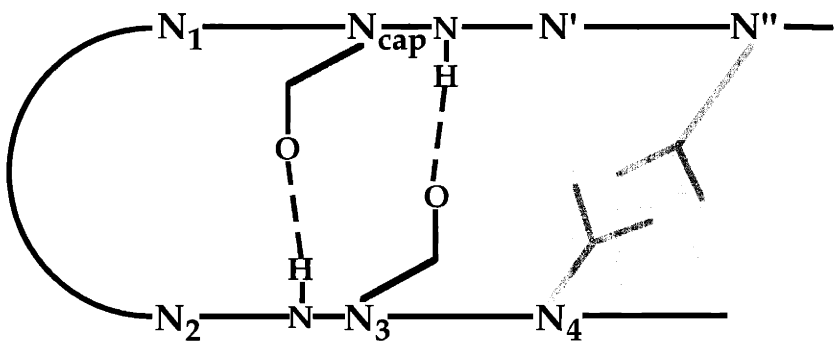
A. N-capping box



C. Schellman motif



B. Big box



D. α_L motif

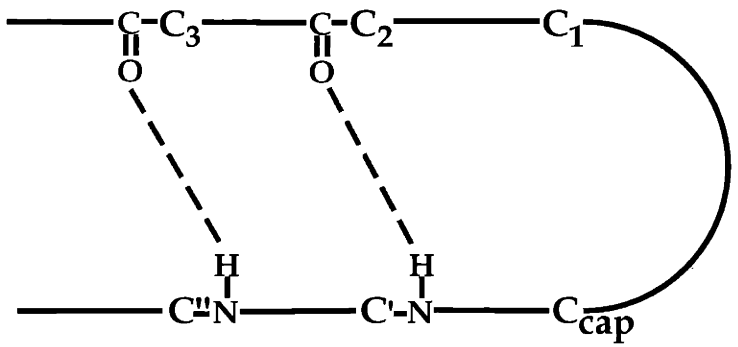


Figure 1. The most common N-terminal motifs are the N-capping box (**A**) and the big box (**B**), whereas the most common C-terminal motifs are the Schellman motif (**C**) and αL motif (**D**). Hydrogen bonds are represented by a dashed, green line. Hydrophobic interactions are represented by a yellow box. (35)

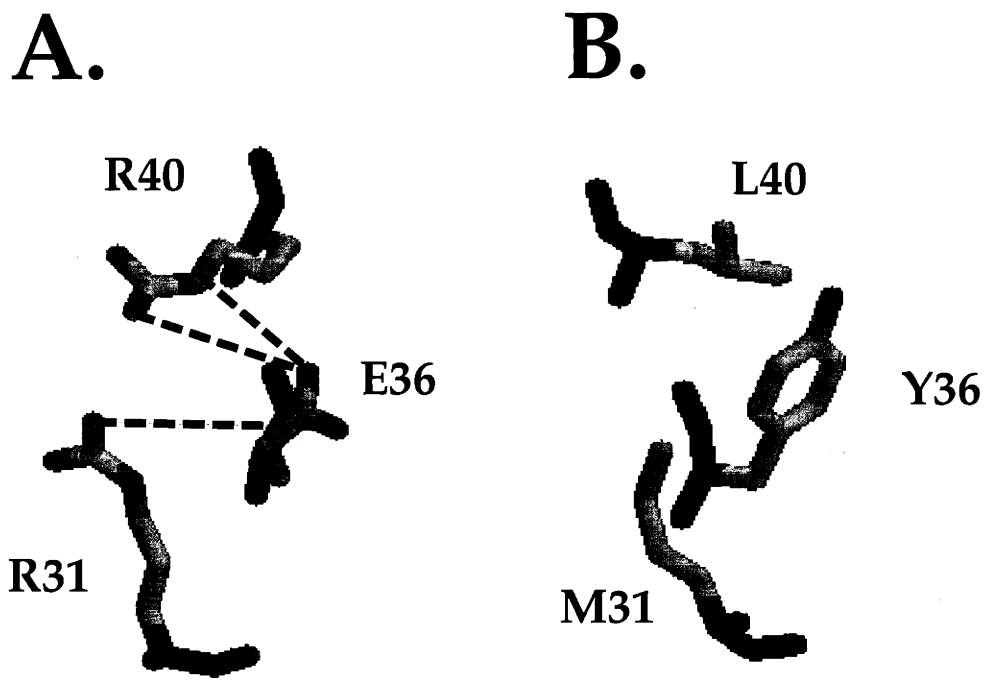


Figure 2. The structure of residues involved in a buried salt-bridge in the wild-type P22 Arc repressor (**A**) and of mutant hydrophobic residues in the Arc MYL variant (**B**). (124, 125)

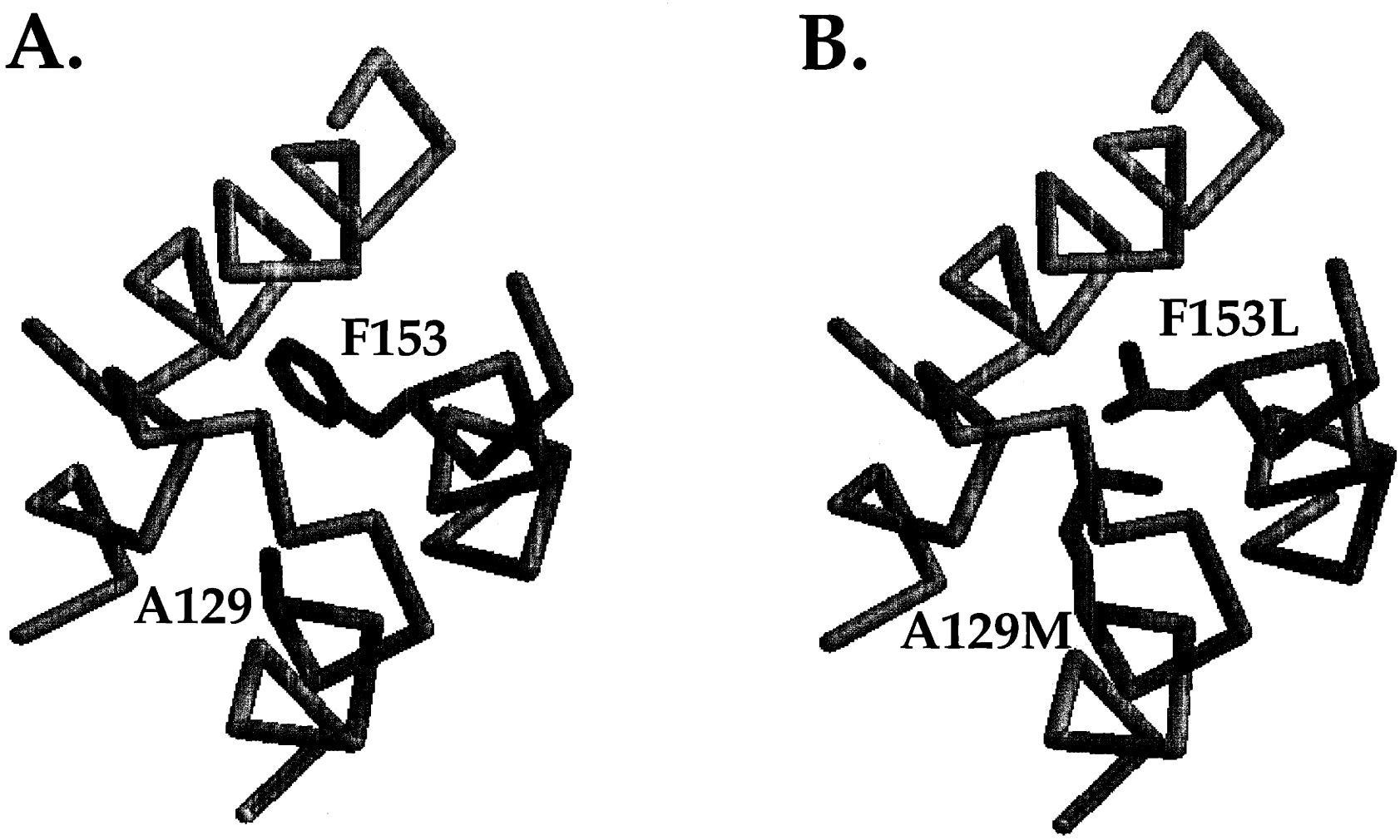


Figure 3. Portions of the structures of wild-type T4 lysozyme (**A**) and the T4 lysozyme double mutant A129M, F153L (**B**). Residues 129 and 153 are labelled in green and their side chains shown in both structures. (126, 127)

Table 1. *Comparison of Secondary-Structure Propensity Results from a Statistical Survey and Mutagenesis Study in Each Element of Secondary Structure.*

		statistical survey[a]	**mutagenesis studies**[b, c, d]
α-helix	**best**	Met, Glu, Leu	Ala, Leu, Met
	worst	Pro, Gly, Ser	Pro, Gly, Asn
β-sheet	**best**	Val, Ile, Phe	Tyr, Thr, Ile
	worst	Pro, Asp, Cys	Pro, Gly, Asp
β -turn	**best**	Pro, Gly, Asp	Gly, Asp, Asn
	worst	Met, Val, Ile	Lys, Ala, Leu

[a] (6). [b] (13). [c] (7). [d] (21).

Table 2. *PrP point mutations that segregate with inherited prion diseases in humans.*[a]

P102L
P105L
A117V
M129V
N171S
D178N
V180I
T183A
F198S
E200K
R208H
V210I
Q217R
E219K
M232R

[a] (106, 107).

Chapter 2

Determinants of Fold Specificity at Arc's N-terminus

Introduction

What sequence information is most important in determining a protein fold? The answer to this question is critical for correctly predicting structure from sequence, for successful *de novo* design, and for understanding protein folding and misfolding. It has been suggested that the binary pattern of polar (P) and hydrophobic (H) residues in a sequence may be one of the most basic features that determines the fold of a protein, whereas other factors such as hydrogen bonding, core packing, electrostatics, and secondary-structure preferences contribute, in varying degrees, to its stability and detailed three-dimensional structure (1,2). The importance of binary pattern has been impressively demonstrated by the isolation of stable α-helical or β-sheet proteins from designed libraries (3-9). Some binary patterns, however, do not uniquely distinguish one protein fold from competing folds. For example, dramatic protein misfolding in prion disease can result from the mutation of one hydrophobic residue to another (10-12).

In the Arc repressor homodimer, interchange of adjacent polar (Asn11) and hydrophobic (Leu12) residues causes a dramatic change in local structure (13). As shown in Figure 1, the surrounding N-terminal sequence folds as an antiparallel, two-stranded β-sheet in the wild-type dimer (14,15) but forms two right-handed 3_{10}-helices in the NL11/LN12 mutant dimer (13), which is also called "switch" Arc. Asn11 is on the protein surface of wild-type Arc, and Leu12 is buried in the core. In "switch" Arc, by contrast, Leu11 is a hydrophobic-core residue, and Asn12 is exposed to solvent. Residues 9-14 have an alternating PHPHPH β-sheet pattern in wild-type Arc and a helical PHHPPH pattern in "switch" Arc. As a result, it is plausible that the binary-

pattern change caused by the NL11/LN12 mutations is largely responsible for the observed sheet-to-helix structural change. However, other explanations are also possible. For example, it might be argued that the precise packing of Leu12 in the protein core specifies the wild-type fold, whereas the detailed packing interactions of Leu11 determine the "switch" Arc fold. In this chapter, I show that Arc mutants containing Asn12 in combination with Gly11, Ala11, Val11, Ile11, Met11, Phe11, or Tyr11 also adopt the "switch" fold. In this instance, therefore, binary pattern appears to be more important than core packing in determining the protein fold.

What happens when binary pattern is insufficient to determine a protein fold uniquely? Arc bearing just the NL11 mutation has an ambiguous binary pattern (PHHHPH) and has been shown to exist in a dynamic equilibrium between the wild-type β-sheet fold (≈ 30% of the molecules) and the "switch" helical fold (≈ 70% of the molecules) under standard solution conditions (16). I found that Arc mutants containing the single Gly11, Ala11, Val11, Ile11, Met11, Phe11, or Tyr11 mutations were also able to access the wild-type and "switch" structures. In this case, however, the identity and β-sheet propensity of residue 11 played a significant role in determining the relative proportions of the two competing protein folds. Finally, I tested the effects of the Gly11, Ala11, Val11, Ile11, Met11, Phe11, and Tyr11 side chains in a mutant background (PL8) that stabilizes the wild-type fold. These mutant variants had properties expected for the wild-type fold but displayed a range of stabilities suggestive of different packing interactions on the surface of the β-sheet.

Materials and Methods

Protein Expression and Purification: Mutations were constructed in the *arc-st11* gene of plasmid pET800 or pSA700 by cassette mutagenesis or Stratagene QuikChange site-directed mutagenesis. The *st11* tag encodes the C-terminal sequence H_6KNQHE (17), which allows affinity purification and improves expression by reducing degradation. Mutants in pET800 or pSA700 backgrounds were overexpressed in *E. coli* strains X90(λDE3) and X90, respectively. Strains used for protein purification were freshly transformed and grown on LB agar plates containing 100 μg/ml ampicillin overnight at 37 °C. A single colony was picked into LB broth plus 100 μg/ml ampicillin and grown for 10-14 hours at 37 °C. This culture was diluted 1:100 into the same medium and growth was continued. When the OD_{600} reached 0.6 to 1.0, the culture was induced with 0.1 mg/ml IPTG and grown for an additional 3.5 hours. Cells were harvested by centrifugation, the supernatant was discarded, and the cell pellet was frozen at -80 °C prior to protein purification.

Mutant proteins were purified by nickel-affinity and anion-exchange chromatography using SP-Sephadex (17). Cells from one liter of cell culture were resuspended in 20 ml of buffer A (0.01 M Tris, 0.1 M NaH_2PO_4, 6 M guanidine hydrochloride, pH 8.0), stirred at 4 °C for 1 hour to allow lysis, and centrifuged at 12,000 X g for 30 minutes. Imidazole was added to the supernatant to a final concentration of 10 mM, and this material was passed over a gravity fed Ni-NTA column (2-3 ml bed size) that had been pre-equilibrated with buffer A plus 10 mM imidazole. Two 50 ml washes with buffer A plus 10 mM imidazole were performed, and bound protein was eluted from

the column using 15 ml of buffer F (0.2 M acetic acid, 6 M guanidine hydrochloride). The eluant was dialyzed against two changes (6 liters each) of buffer B (0.01 M Tris, 0.2 mM EDTA, pH 7.5) for a minimum of 8 hours. The protein sample was then loaded onto a gravity fed SP-Sephadex column (2-3 ml bed size) that had been pre-equilibrated with buffer B plus 100 mM KCl. This column was washed with 100 ml of buffer B plus 100 mM KCl and eluted with buffer B plus 2 M KCl. Fractions containing the purified protein were dialyzed against two 6-liter changes of storage buffer (50 mM Tris, 250 mM KCl, 0.2 mM EDTA, pH 7.5) for a minimum of 8 hours. For ^{15}N-labelling of mutants for NMR, proteins were overexpressed from pET800 in *E. coli* strain BL21(λDE3)-pLysS using M9T minimal medium containing $^{15}NH_4Cl$ (0.8 g/liter) as the sole nitrogen source. The ^{15}N-labelled mutants were purified in the manner described above, but after purification were dialyzed against three changes of 6 liters of distilled H_2O for a minimum of 8 hours and then lyophilized.

Circular Dichroism and Fluorescence: Circular-dichroism (CD) spectra were taken in an AVIV Circular Dichroism Spectrometer Model 60DS as described (18). Far-ultraviolet CD spectra from 200 nm to 250 nm (step size 0.5 nm; averaging time 10 seconds) were determined at 10 μM protein concentration in storage buffer at 15 °C using a 1 cm pathlength cuvette. Near-ultraviolet CD spectra from 260 nm to 300 (step size 0.5 nm; averaging time 3 seconds) were the average of 5 scans determined at 100 μM protein concentration in storage buffer at 15 °C in a 1 cm pathlength cuvette. CD spectra were baseline corrected using the far-ultraviolet and near-ultraviolet CD spectra of storage buffer. Fluorescence emission spectra from 300 to 420 nM were determined in a Photon

Technology International QM-2000-4SE instrument using an excitation wavelength of 280 nm. Spectra were the average of 3 scans (step size 1 nm; averaging time 1 second) determined using 100 μM protein in storage buffer at 15 °C in a 1 cm pathlength cuvette.

Nuclear Magnetic Resonance: Lyophilized ^{15}N-labelled protein was resuspended at a concentration of ~4 mM in a buffer containing 20 mM $NaPO_4$ (pH 5.0), 10% D_2O, and 1 mM 3-(trimethylsilyl)-propionic acid (TMSP) as the internal chemical shift standard. The pH was adjusted to 4.9 using HCl, but corrections for the effect of D_2O on the measured pH were not made (19). NMR spectra at 30 °C were acquired using a Bruker DMX500 spectrometer. The ^{1}H and ^{15}N chemical shifts are relative to TMSP (20). Heteronuclear single quantum coherence spectra (HSQC) were acquired with standard Bruker pulse sequences (21), and pulsed-field gradients were used for coherence selection and solvent suppression (22). Presaturation was used for solvent suppression for the wild-type sample. For the mutants, data were collected as the average of 4 scans of 2048 data points for each of 256 t1 transients. For "switch" Arc, the data were collected with 8 scans of 2048 data points for each of 64 t1 transients. The raw data were apodized with shifted sine bell functions and zero filled in both dimensions before Fourier transformation using the Bruker Xwin-NMR software.

Denaturation Assays: Thermal and chemical denaturation experiments were performed and analyzed as described using 1 cm pathlength cuvettes (17,23). For thermal denaturation studies, CD ellipticity at 222 nm was monitored from 4 to 100 °C, with a step size of 2 °C, an averaging time of 30 seconds, and an equilibration time of 1 minute.

The protein concentration was 10 μM in storage buffer. For chemical denaturation studies, ellipticity at 230 nm was monitored at 25 °C for samples containing from 0 to 9 M urea. At each urea concentration, the ellipticity of 5 μM protein in storage buffer was averaged for 60 data points taken at 1 second intervals following 1 minute of sample mixing.

Denaturation curves were fitted using a nonlinear least-squares procedure implemented in the program NonLin (24). Arc dimers denature in a concerted reaction to two unfolded monomers (25). For thermal denaturation studies, curves were fitted using the following equations (17):

$$\Delta G^\circ_u = \Delta H^\circ_u - (T / T_m) * [\Delta H^\circ_u + R * T_m * ln(P_t)] + \Delta C^\circ_p * [T - T_m - T * ln(T / T_m)]$$

$$K_u = exp(-\Delta G^\circ_u / R * T)$$

$$f_u = [-K_u + (K_u^2 + 8 * P_t * K_u)^{1/2}] / (4 * P_t)$$

$$\varepsilon_{obs} = f_u * (\varepsilon_D + ds * T) + (1 - f_u) * (\varepsilon_N + ns * T)$$

For chemical denaturation studies, curves were fitted using the following equations:

$$K_u = exp((-\Delta G^\circ_u - m * [urea])) / (R * T))$$

$$f_u = [-K_u + (K_u^2 + 8 * P_t * K_u)^{1/2}] / (4 * P_t)$$

$$\varepsilon_{obs} = f_u * (\varepsilon_D + ds * [urea]) + (1 - f_u) * (\varepsilon_N + ns * [urea])$$

where ε_{obs} is the observed ellipticity; ε_D and ε_N are the intercept ellipticities of the denatured and native proteins; *ds* and *ns* are the denatured and native baseline slopes; f_u is the fraction of protein unfolded; K_u is the equilibrium constant for Arc unfolding and dissociation; ΔG°_u, ΔH°_u, and ΔC°_p are the free energy, enthalpy, and heat capacity

changes from unfolding and dissociation; T is the temperature in Kelvin; T_m is the temperature in Kelvin where 50% of the protein is unfolded; $[urea]$ is the urea concentration; R is the universal gas constant; and P_t is the Arc concentration in monomer equivalents. For each thermal denaturation curve, ΔC°_p was fixed at 1.31 kcal/mol (23) and ΔH°_u, T_m, and the slopes and intercepts of the denatured and native baselines were calculated from fitting.

Results

Mutant effects in the Asn12 (LN12) background: In "switch" Arc, the Leu11 side chain is part of the hydrophobic core and the Asn12 side chain is on the protein surface (13). Will any non-polar residue at position 11 also result in a structure similar to "switch" Arc if the Asn12 substitution is also present? To address this question, I constructed, purified, and characterized Arc variants containing Gly, Ala, Val, Ile, Met, Phe, or Tyr at position 11 in an Asn12 mutant background. Each of these proteins had a far-UV CD spectrum expected for a molecule with significant secondary structure and showed a cooperative thermal melt with a T_m between 46 and 59 °C (Table 1; Fig. 2A and 2B show spectra and melting curves for selected mutants). Hence, like "switch" Arc, other position 11 mutants in the Asn12 background also fold into a reasonably stable tertiary structure.

As noted previously (13), the fluorescence spectra of "switch" Arc was red shifted and had higher intensity than the spectra of wild-type Arc (Fig. 3A). In the cases tested, the other position 11 mutants in the Asn12 background also had fluorescence spectra

more similar to "switch" Arc than to wild-type Arc (Fig. 3A). The near-UV CD spectra of the "switch" and wild-type structures (Fig. 3B) were also distinct, with the latter protein showing positive ellipticity from 260-275 nm and distinct negative minima at 285 and 292 nm (13). As shown for six mutants in Fig. 3B, the other position 11 mutants in the Asn12 background had near-UV CD spectra more similar to "switch" Arc than to wild-type Arc. For example, none of the Gly11, Ala11, Val11, Ile11, Met11, Phe11, or Tyr11 substitutions displayed negative minima at 285 and 292 nm in the Asn12 mutant background. As a final comparison, I collected two-dimensional HSQC NMR spectra for the Ala11/Asn12, Ile11/Asn12, and Met11/Asn12 substituted proteins. These spectra displayed a number of resonances that are characteristic of the 3_{10}-helical region of "switch" Arc (circled resonances in Fig. 4) but are absent in the wild-type Arc spectrum. Taken together, these spectroscopic studies suggest that the Gly11/Asn12, Ala11/Asn12, Val11/Asn12, Ile11/Asn12, Met11/Asn12, Phe11/Asn12, and Tyr11/Asn12 substituted Arc variants assume structures similar to the structure of the Leu11/Asn12 or "switch" Arc molecule.

Structural preferences in the context of an ambiguous binary pattern: Cordes *et al.* (16) showed that Arc bearing just the Leu11 mutation adopts the wild-type structure roughly 30% of the time and the "switch" structure roughly 70% of the time, as estimated from near-UV spectra taken in storage buffer at 15 °C or fluorescence spectra taken in storage buffer at 25 °C. I constructed and purified singly mutated Arc variants containing Gly11, Ala11, Val11, Ile11, Met11, Phe11, and Tyr11. These mutants displayed cooperative melts with T_m*'s* between 56 and 71 °C (Table 1; Fig. 5). Based upon their near-UV and

fluorescence spectra, different non-polar residues at position 11 favored the β-sheet and helical folds to varying degrees (Fig. 6; Table 2). The fraction of Arc position-11 mutants in the wild-type (sheet) fold was determined by fitting their near-UV CD spectra using the near-UV CD spectra of the same position-11 mutants in the Asn12 and Leu8 backgrounds as the helical and sheet basis spectra (Fig. 7). The Leu11, Ala11, and Gly11 single mutants had the highest apparent equilibrium population of the "switch" fold, whereas the single Ile11 and Val11 mutants had spectra very similar to wild-type Arc. These results suggest that β-branched residues at position 11 strongly favor the wild-type β-sheet fold. The single mutants containing Met11, Phe11, or Tyr11 showed more "switch" character than the wild type, Ile11 or Val11 variants, but displayed less "switch" character than Leu11, Ala11, or Gly11.

The equilibrium populations of the helical and β-sheet forms of Arc NL11 were previously shown to be strongly influenced by solvent composition and temperature (16). For example, at 0 °C in buffer containing 250 mM KSCN, Arc NL11 almost exclusively adopted the helical switch fold, whereas at room temperature in buffer containing 10% ethanol, this mutant existed largely in the wild-type β-sheet conformation. Altering temperature and solvent composition also affected the near-UV spectra of several of the other position 11 mutants. At 0 °C in buffer containing KSCN, the far-UV CD spectrum of the NG11, NM11, NI11, NF11, and NY11 mutants were shifted to positions expected for an increased equilibrium population of the helical structure relative to the sheet structure (spectra from selected mutants are shown in Fig. 8). At 15 °C in 10% ethanol, by contrast, the NG11, NM11, NF11, and NY11 mutant spectra indicated an increase in

the β-sheet form relative to the helical form (Fig. 8). These results suggest that these position-11 mutants are able to adopt both alternative structures, with the equilibrium between these structures being strongly influenced by the chemical identity of the mutant side chain as well as by temperature and solvent composition.

Position 11 effects in the Arc PL8 background: The Arc PL8 mutation results in formation of an additional hydrogen bond at each end of the β-sheet, creating a hyperstable mutant with a structure nearly identical to wild-type Arc (26). I hypothesized that the PL8 substitution should stabilize mutants with nonpolar substitutions at position 11 in the wild-type conformation, thereby eliminating significant equilibrium levels of the competing helical structure. Mutants containing the PL8 substitution and the NG11, NA11, NI11, NL11, NM11, or NY11 substitutions were constructed and purified. The fluorescence spectra of wild-type Arc, Arc PL8, and the position-11 mutants in the PL8 background were similar (Fig. 9A), as were their near-UV CD spectra (Fig. 9B). These results provide further evidence that the position-11 mutations can adopt an essentially wild-type structure. The T_m*'s* of the position 11 mutants in the PL8 background varied from 72 to 80 °C (Table 1), with the Ile11 variant having the highest stability.

Do stabilities in different backgrounds explain sheet/helical preferences: Ile11 clearly favored the β-sheet conformation in an otherwise wild-type context, whereas Leu11 favored the "switch" conformation. The net balance between the favorable and unfavorable interactions made by each side chain in each structural context ultimately determines structural preference. Thus, Ile11 must make better packing interactions on

the surface of the β-sheet than in the core of the "switch" structure, whereas the opposite is true of Leu11. In some cases, the structural preferences observed for the position 11 mutants in otherwise wild-type backgrounds were also evident, albeit subtly, in terms of the relative thermal stabilities of the mutants in the LN12 and PL8 backgrounds. For example, the T_m of Ile11 variant was 1 degree less than the Leu11 variant in the LN12 background ("switch" conformation) and 4 degrees higher in the PL8 background (wild-type conformation) (Table 1). In other cases, however, the relative thermal stabilities in the LN12 and PL8 backgrounds failed to predict the expected structural preference. For example, the T_m of Met11 variant was 2 degrees lower than the Leu11 variant in the LN12 background and 3 degrees lower in the PL8 background (Table 1). On this basis, the Met11 mutant would have been expected to have more "switch" character than the Leu11 mutant, but the opposite was observed (Table 2).

Because of differences in the enthalpies or heat capacities of thermal denaturation, the relative stabilities of some of the position 11 mutants in the LN12 and PL8 backgrounds could be different at 15 °C, the temperature where structural preferences were determined. Alternatively, the temperature dependence of equilibrium between the wild-type and "switch" conformations may change the structural preferences at higher temperatures (16). In an attempt to obviate this problem, I also determined the stabilities to urea denaturation at 25 °C of mutants with Ala11, Leu11, Ile11, Met11, and Phe11 in the wild-type, LN12, and PL8 backgrounds (Fig. 10; Table 3). Again, in some cases, the relative stabilities in the LN12 and PL8 backgrounds predict the observed structural preferences (cf. Ile11 and Leu11), whereas, in other cases, they do not (cf. Met11 and

Leu11). It appears that stabilities in the LN12 and PL8 backgrounds are not a reliable predictor of structural preference in an otherwise wild-type background. This may occur because of mutant effects on the unfolded state.

Discussion

Previous studies have suggested that binary pattern is the key determinant of the gross three-dimensional fold of a protein whereas other interactions define the detailed aspects of the structure and stability of the native protein fold (3-5,7-9,13,27-29). The results presented here are generally consistent with this idea with one important additional stipulation. Namely, if competing structures are accessible to a protein because of an ambiguous binary pattern, then the detailed interactions mediated by one or a few side chains can determine the dominant conformation. For example, when residues 9-14 of Arc repressor had the ambiguous pattern PH**H**HPH, the identity of residue 11 was critical for determining the relative proportions of the wild-type fold (antiparallel β-sheet) and the "switch" fold (two 3_{10} helices). When position 11 was isoleucine, almost all molecules assumed the β-sheet conformation. By contrast, when position 11 was leucine, roughly 70% of the molecules adopted the "switch" fold. Hence, in this ambiguous background, moving a single methyl group from the β-carbon to the γ-carbon of the position-11 side chain results in a dramatic difference in structural preference.

The "switch" Arc fold was originally characterized in a mutant protein containing the NL11 and LN12 mutations (13). Cordes et al. (13) proposed that the change from the wild-type to the mutant structure resulted largely from the change from a β-sheet pattern

(PHPHPH) for residues 9-14 in wild-type Arc to a helical pattern (PH**HP**PH) in "switch" Arc. In the "switch" structure, Asn12 was a surface residue, whereas Leu11 was buried in the protein core. In the Asn12 background, I found that variants with almost any non-polar residue at position 11 folded into the "switch" conformation as assayed by near-UV CD, fluorescence, and NMR. These mutants had different thermal stabilities; the Gly11/Asn12 mutant (T_m = 46 °C) was least stable and the Phe11/Asn12 mutants (T_m = 59 °C) was most stable. The best correlations between T_m's and side-chain properties for these mutants were with hydrophobicity (R = 0.93) and molecular weight (R=0.88) (Fig. 11). The correlations between T_m's and secondary-structure preferences for α-helices (R = 0.36) or β-sheets (R = 0.41) were poor. The fact that the "switch" fold can be formed with almost any non-polar residue at position 11 in the Asn12 background is consistent with the idea that the change in binary pattern rather than the detailed properties of the position-11 side chain cause the observed structural rearrangement.

Different non-polar residues at position 11 were also compatible with the wild-type Arc fold, in which they would occupy surface positions. In the PL8 (Leu8) background, which stabilizes the wild-type structure (23,26,30), all of the position-11 variants had near-UV CD and fluorescence spectra similar to those of wild-type Arc. These mutants also showed a range of stabilities, with Gly11/Pro8 being the least stable (T_m = 70 °C) and Ile11/Pro8 being the most stable (T_m = 80 °C). The T_m's of these variants showed only a modest correlation with hydrophobicity (R = 0.55) and poorer correlations with molecular weight (R = 0.44), β-sheet propensity (R = 0.31), α-helical propensity (R = 0.26), and molecular weight (R = 0.44) (Fig. 11).

In the wild-type Arc dimer, Asn11 and Asn11' are surface exposed and immediately adjacent to each other across the antiparallel β-sheet. Although the main-chain atoms of these residues are not hydrogen bonded, the side chain of Asn11 in one subunit hydrogen bonds with the side chain of Asn11' in other subunit (14). In the wild-type conformation, hydrophobic residues at position 11 would also be positioned to pack specifically against each other. Indeed, strong statistical correlations exist between specific cross-strand residue pairs in antiparallel β-sheets, as would be expected if certain side chains at these positions interact closely and favorably (31,32). In the B1 domain of protein G, mutagenesis studies have shown that cross-strand Ile-Ile surface packing was more favorable than Ala-Ala packing by roughly 2 kcal/mol (32), and a similar difference (1.7 kcal/mol) is observed between Ile11-Ile11' and Ala11-Ala11' in the Pro8 mutant background. Hence, it seems likely that a major source of the stability differences between position-11 variants in this background arises from differences in hydrophobic interactions and in detailed packing.

In an otherwise wild-type Arc background, the different position-11 variants adopted the β-sheet and 3_{10}-helical conformations to varying degrees. As shown in Table 2, the equilibrium populations of the two competing structures in storage buffer at 15 °C ranged from greater than 95% β-sheet to about 70% 3_{10}-helix. I found a strong correlation (R = 0.90) between the fraction of molecules in the β-sheet conformation and the β-sheet propensities of the residue-11 side chain (Fig. 12). Correlations with α-helical propensity (R = 0.15), molecular weight (R = 0.40), and hydrophobicity (R =

0.58) were much weaker (not shown). Because intrinsic β-sheet propensities must reflect the ability of a side chain to make favorable interactions or avoid unfavorable interactions in the context of a β-sheet relative to other conformations, it appears that this factor may be most important in determining the fold preference when both structures are accessible because of the ambiguous binary pattern.

Irrespective of what factors determine the structural preference, it is clear that the majority of hydrophobic residues favor the β-sheet or wild-type fold when placed at position 11 in otherwise wild-type Arc. Has the Arc fold evolved to make it relatively resistant to changing structure when hydrophobic residues are present on the surface of the β-sheet? The surface residues of this sheet interact with bases in the major groove of operator DNA, thereby determining operator-binding specificity (15). As a result, placing a non-polar side chain at position 11 might be required in some instances for recognition of certain base sequences. Indeed, non-polar residues—including Gly, Val, Met, and Leu—are observed at the position corresponding to Asn11 of Arc in a number of paralogs (Fig. 13).

There are a number of interesting experiments for the future. It would be useful to determine the structure of the Ile11 or Val11 mutant to verify how these side chains stabilize the wild-type fold. Arc is small enough to make by peptide synthesis, allowing non-natural amino acids to be introduced into the β-sheet (33). This would allow exploration of the effects of a larger repertoire of residues of diverse size, branching, and polarity on the sheet/helix equilibrium. One could also place Ile or Val at position 11 and

then attempt to stabilize the "switch" fold in ways other than by introducing Asn at position 12. If binary pattern is really the dominant determinant of structural choice, then any polar residue at position 12 should be equally effective at promoting the "switch" conformation. In an Ile11-Leu12 background, one could try to redesign the hydrophobic cores of both conformations to stabilize the "switch" conformation and destabilize the wild-type conformation. Alternatively, one could begin with a Leu11-Leu12 background and attempt to stabilize the β-sheet conformation in ways other than introduction of Leu8. Such experiments would allow very sensitive tests of our ability to design protein structures.

References

1. Sauer, R. T. (1996) *Fold Des* **1**(2), R27-30

2. Cordes, M. H., Davidson, A. R., and Sauer, R. T. (1996) *Curr Opin Struct Biol* **6**(1), 3-10.

3. Weigent, D. A., Clarke, B. L., and Blalock, J. E. (1994) *Immunomethods* **5**(2), 91-7.

4. Yan, Y., and Erickson, B. W. (1994) *Protein Sci* **3**(7), 1069-73.

5. Xu, G., Wang, W., Groves, J. T., and Hecht, M. H. (2001) *Proc Natl Acad Sci U S A* **98**(7), 3652-7.

6. Rojas, N. R., Kamtekar, S., Simons, C. T., McLean, J. E., Vogel, K. M., Spiro, T. G., Farid, R. S., and Hecht, M. H. (1997) *Protein Sci* **6**(12), 2512-24.

7. Roy, S., and Hecht, M. H. (2000) *Biochemistry* **39**(16), 4603-7.

8. Marshall, S. A., and Mayo, S. L. (2001) *J Mol Biol* **305**(3), 619-31.

9. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., and Hecht, M. H. (1993) *Science* **262**(5140), 1680-5.

10. Cohen, F. E. (1999) *J Mol Biol* **293**(2), 313-20.

11. Carrell, R. W., and Gooptu, B. (1998) *Curr Opin Struct Biol* **8**(6), 799-809.

12. Prusiner, S. B. (1994) *Philos Trans R Soc Lond B Biol Sci* **343**(1306), 447-63.

13. Cordes, M. H., Walsh, N. P., McKnight, C. J., and Sauer, R. T. (1999) *Science* **284**(5412), 325-8.

14. Breg, J. N., van Opheusden, J. H., Burgering, M. J., Boelens, R., and Kaptein, R. (1990) *Nature* **346**(6284), 586-9.

15. Raumann, B. E., Rould, M. A., Pabo, C. O., and Sauer, R. T. (1994) *Nature* **367**(6465), 754-7.

16. Cordes, M. H., Burton, R. E., Walsh, N. P., McKnight, C. J., and Sauer, R. T. (2000) *Nat Struct Biol* **7**(12), 1129-32.

17. Milla, M. E., Brown, B. M., and Sauer, R. T. (1993) *Protein Sci* **2**(12), 2198-205.

18. Milla, M. E., and Sauer, R. T. (1995) *Biochemistry* **34**(10), 3344-51.

19. Bundi, A., and Wuthrich, K. (1979) *Biopolymers* **18,** 285-297

20. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) *J Biomol NMR* **6**(2), 135-40.

21. Bodenhausen, G., and Ruben, D. J. (1980) *Chemical Physics Letters* **69,** 185-189

22. Davis, A. L., Keeler, J., Laue, E. D., and Moskau, D. (1992) *Journal of Magnetic Resonance* **98,** 207-216

23. Milla, M. E., Brown, B. M., and Sauer, R. T. (1994) *Nat Struct Biol* **1**(8), 518-23.

24. Johnson, M., Frasier, S. (1985) *Methods in Enzymology* **117,** 301-342

25. Bowie, J. U., and Sauer, R. T. (1989) *Proc Natl Acad Sci U S A* **86**(7), 2152-6.

26. Schildbach, J. F., Milla, M. E., Jeffrey, P. D., Raumann, B. E., and Sauer, R. T. (1995) *Biochemistry* **34**(4), 1405-12.

27. Huang, E. S., Subbiah, S., and Levitt, M. (1995) *J Mol Biol* **252**(5), 709-20.

28. Handel, T. M., Williams, S. A., and DeGrado, W. F. (1993) *Science* **261**(5123), 879-85.

29. Xiong, H., Buckwalter, B. L., Shieh, H. M., and Hecht, M. H. (1995) *Proc Natl Acad Sci U S A* **92**(14), 6349-53.

30. Brown, B. M., Milla, M. E., Smith, T. L., and Sauer, R. T. (1994) *Nat Struct Biol* **1**(3), 164-8.

31. Wouters, M. A., and Curmi, P. M. (1995) *Proteins* **22**(2), 119-31.

32. Smith, C. K., and Regan, L. (1995) *Science* **270**(5238), 980-2.

33. Wales, T. E., and Fitzgerald, M. C. (2001) *J Am Chem Soc* **123**(31), 7709-10.

34. Levitt, M. (1978) *Biochemistry* **17**(20), 4277-85.

35. Fauchere, J. L., Charton, M., Kier, L. B., Verloop, A., Pliska, V. (1988) *Int J Pept Protein Res* **32**(4), 269-78.

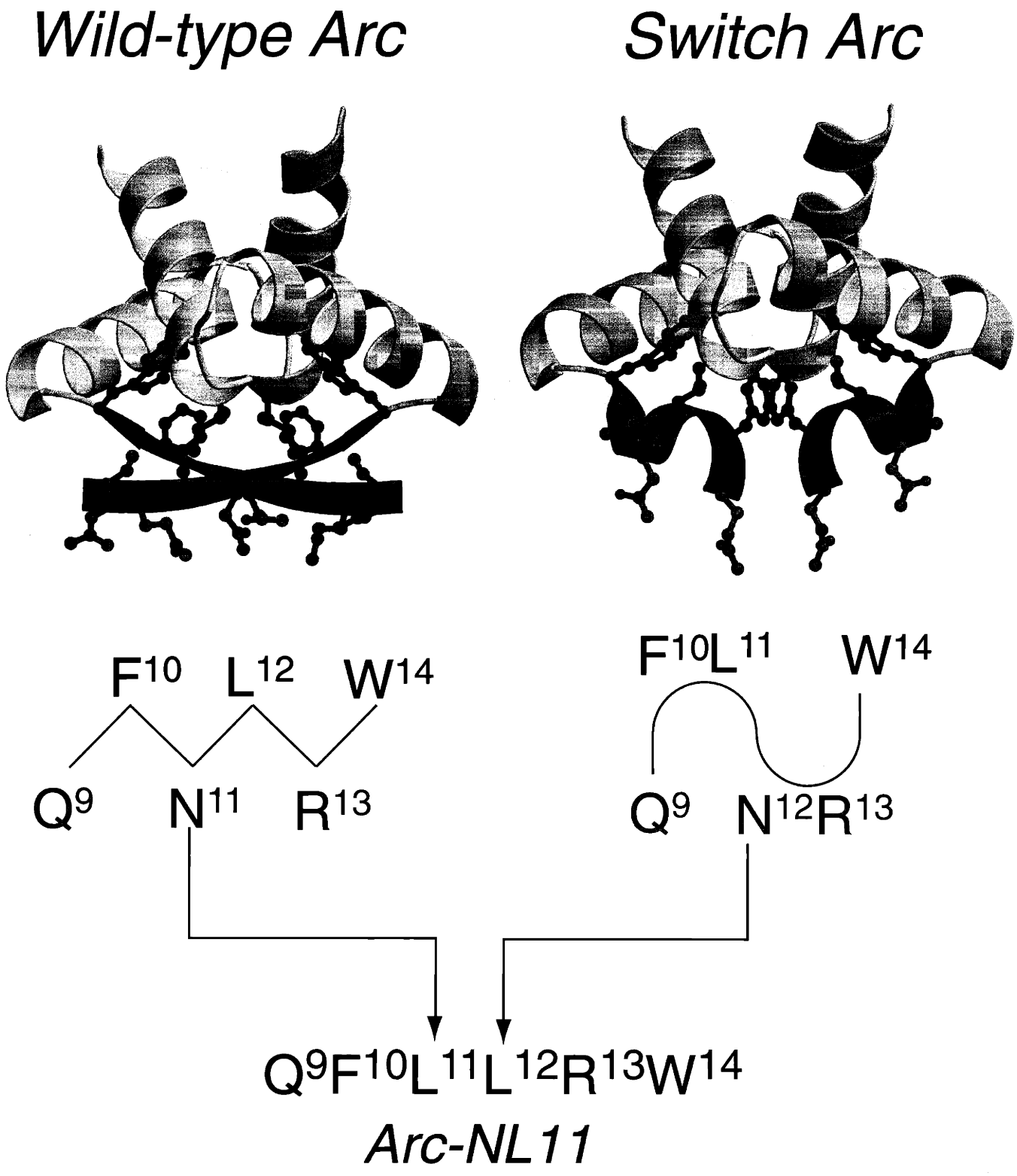


Figure 1. Partioning of hydrophobic and polar residues between the interior and surface in the N-terminal region of wild-type and switch (NL11, LN12) Arc. Buried residues are colored blue and surface residues are colored cyan. The binary pattern of residues 9-14 in wild-type Arc (PHPHPH) is that of a solvent-exposed β-sheet. The switch binary pattern (PHHPPH) is that of a solvent-exposed helix. The NL11 binary pattern is compatible with both folds. (Adapted from Ref. 16)

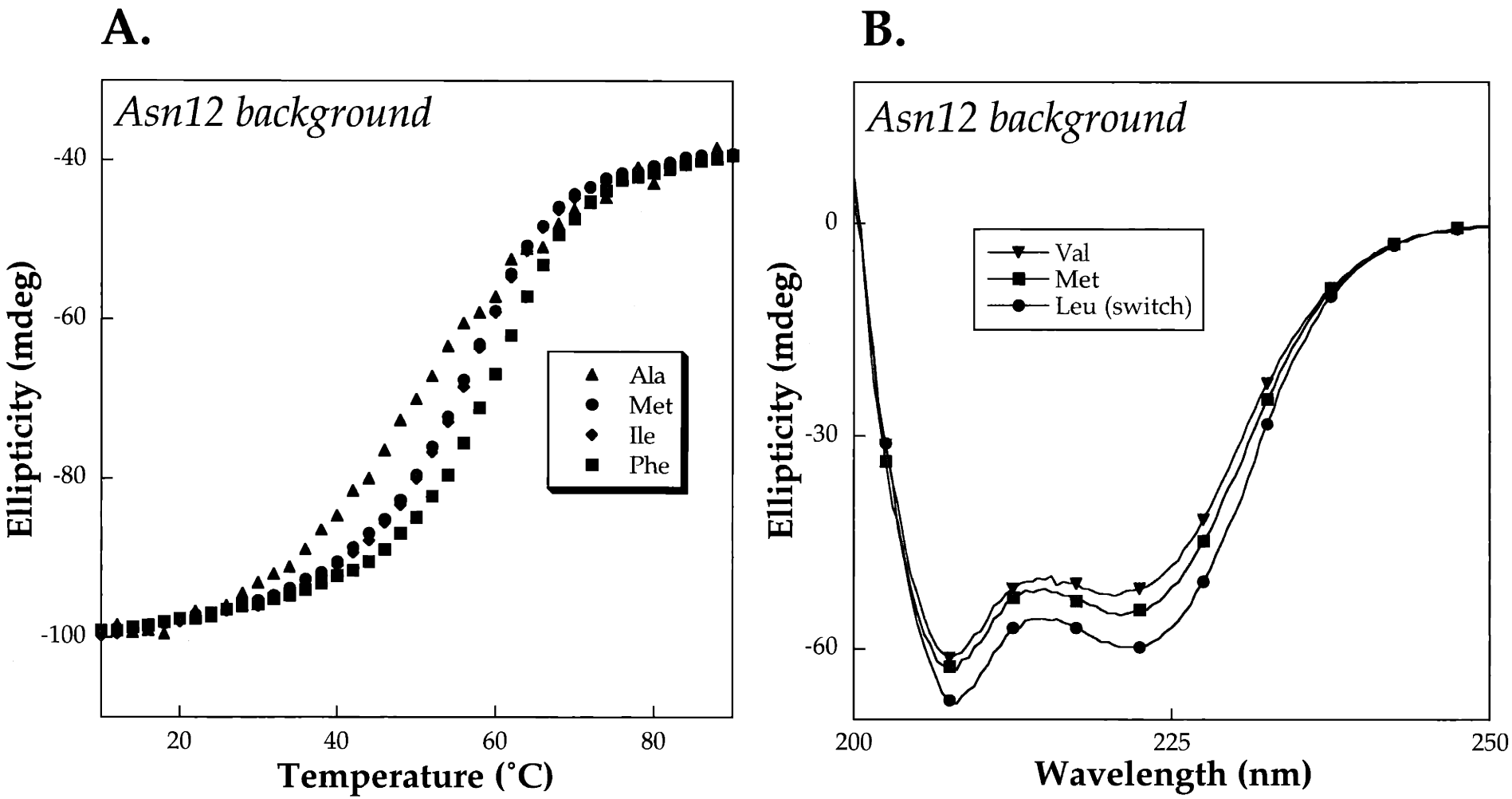


Figure 2. Thermal melts (**A**) and far-UV CD spectra (**B**) of position-11 mutants in the Asn12 background. For all experiments, Arc variants were in a buffer containing 50 mM Tris, 250 mM KCl, 0.2 mM EDTA, pH 7.5. Far-UV CD spectra were taken at 15 °C.

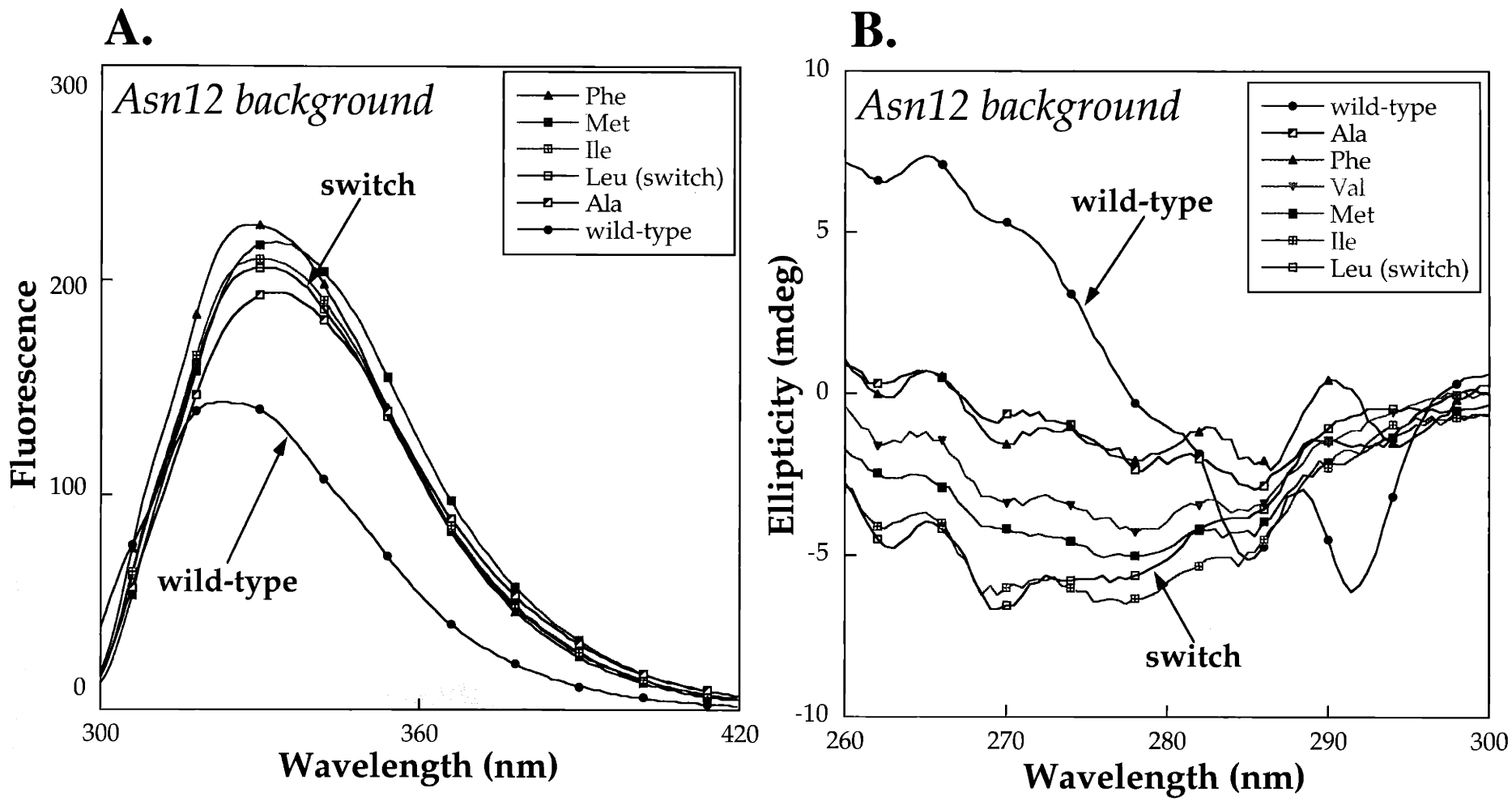


Figure 3. Fluorescence (**A**) and near-UV CD spectra (**B**) of wild-type Arc and position-11 mutants in the Asn12 background. For all experiments, Arc variants were at 100 μM in a buffer containing 50 mM Tris, 250 mM KCl, 0.2 mM EDTA, pH 7.5. Fluorescence spectra were taken at 25 °C. Near-UV CD spectra were taken at 15 °C.

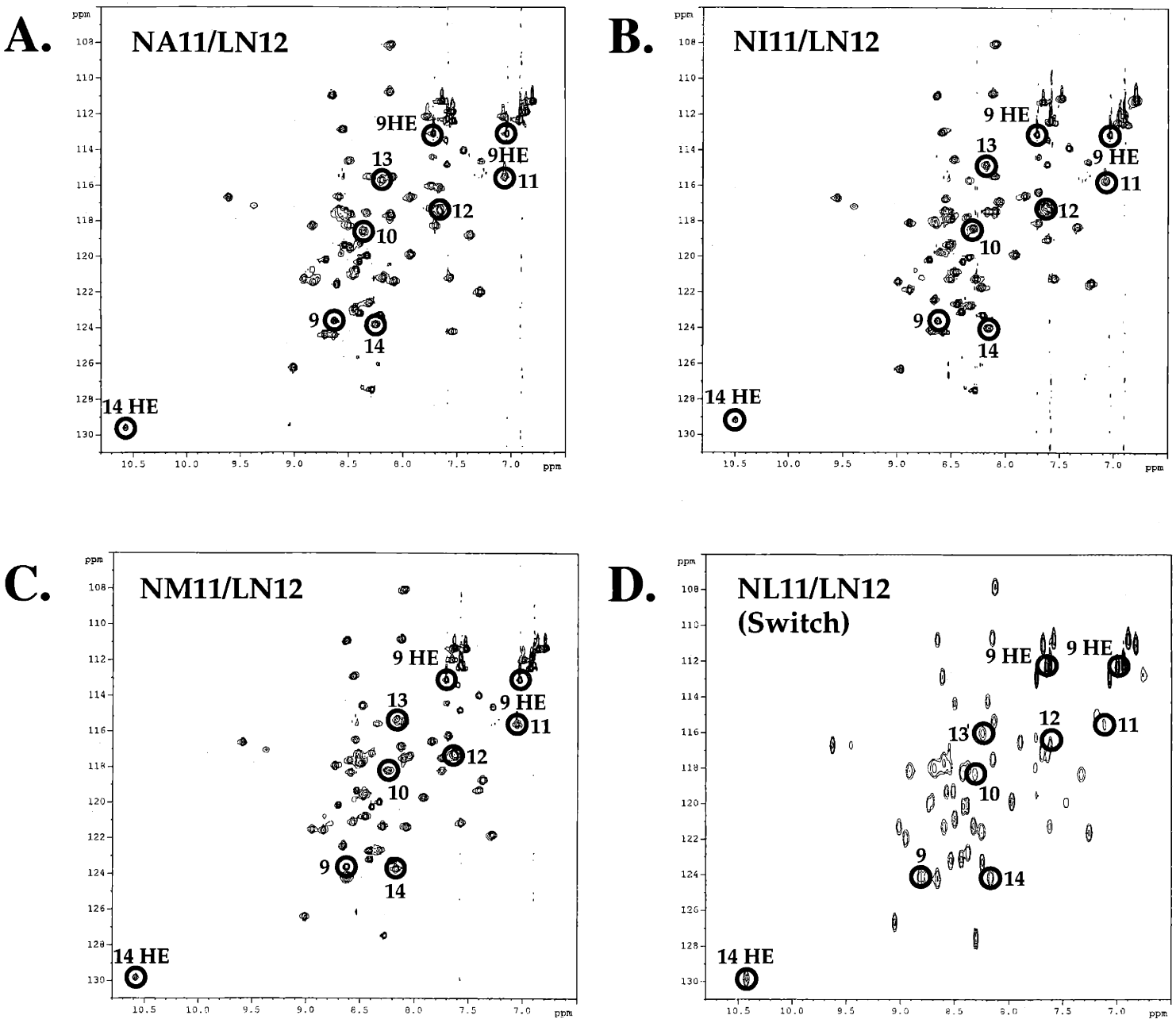


Figure 4. HSQC NMR spectra of position-11 mutants (~4 mM) in the Asn12 background. Resonances assigned to residues 9, 10, 11, 12, 13, and 14 in "switch" Arc are labelled and circled in each spectra. NMR spectra were acquired at 30 °C with samples in a buffer containing 20 mM $NaPO_4$, 10% D_2O, pH 4.9 and 1 mM 3-(trimethylsilyl)-propionic acid (TMSP) as the internal chemical shift standard.

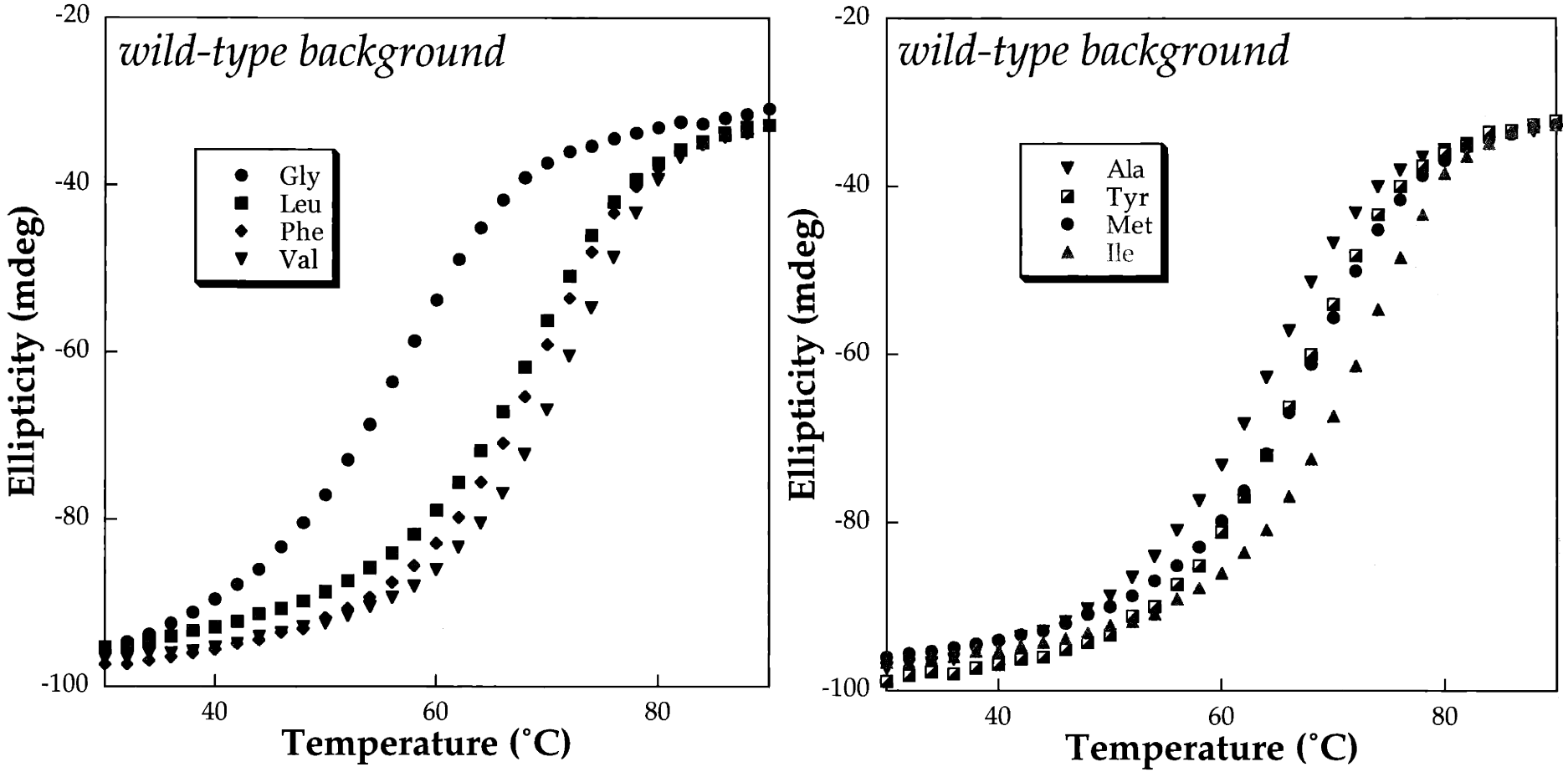


Figure 5. Both panels show thermal denaturation curves for position-11 mutants in the wild-type background. Thermal melts were taken at 10 μM protein concentration in a buffer containing 50 mM Tris, 250 mM KCl, 0.2 mM EDTA, pH 7.5.

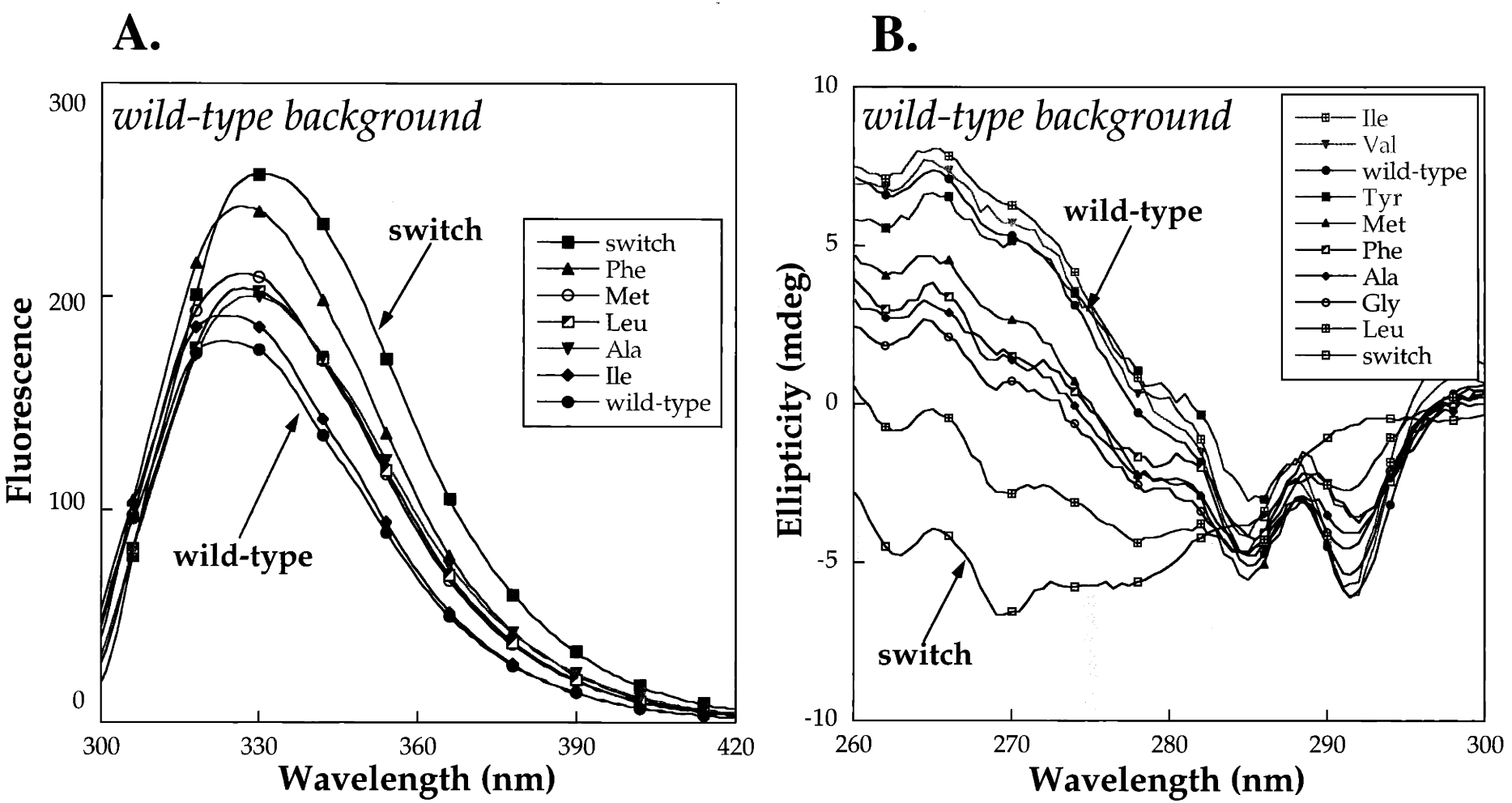


Figure 6. Fluorescence (**A**) and near-UV CD spectra (**B**) of switch Arc, wild-type Arc, and position-11 mutants in the wild-type background. For all experiments, Arc variants were at 100 μM in a buffer containing 50 mM Tris, 250 mM KCl, 0.2 mM EDTA, pH 7.5. Fluorescence spectra were taken at 25 °C. Near-UV CD spectra were taken at 15 °C.

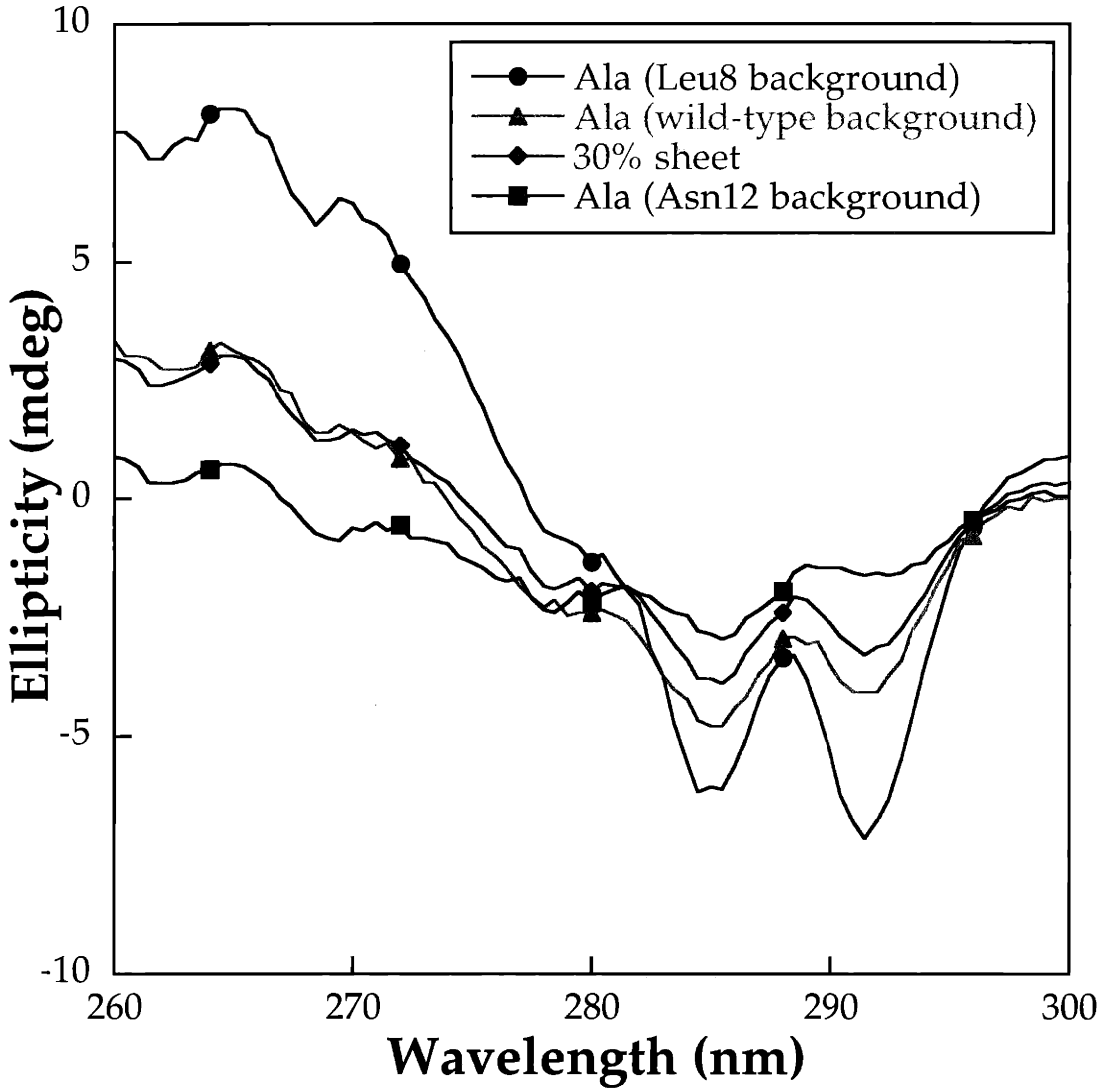
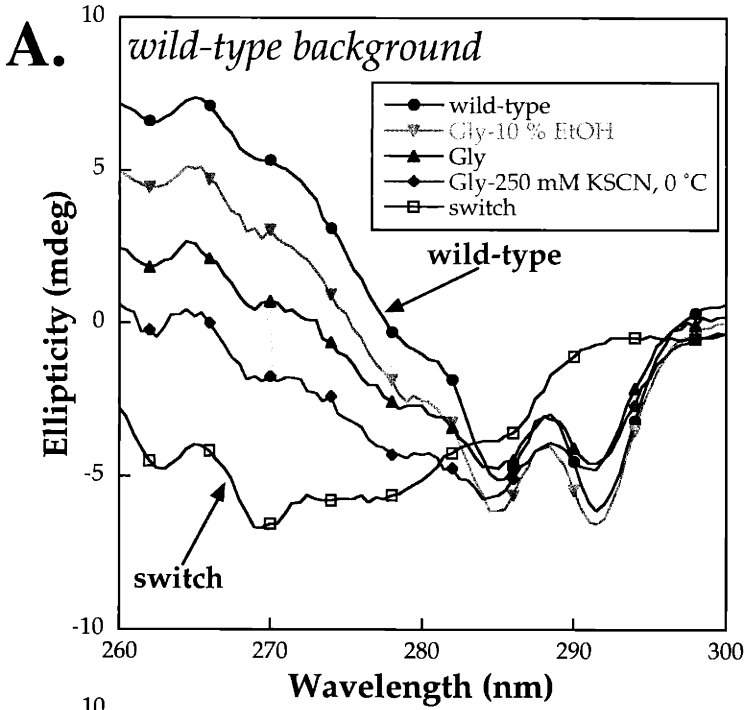
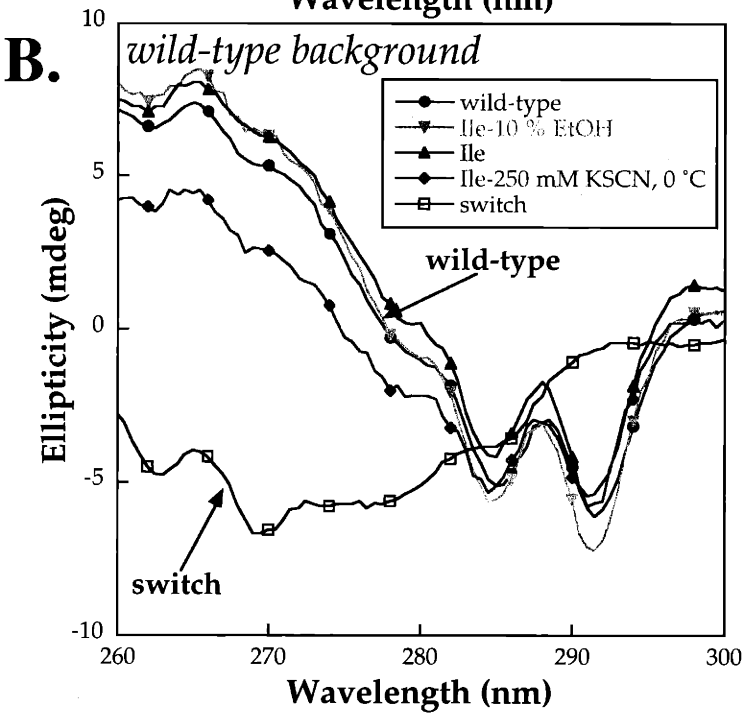


Figure 7. Determination of the fraction of Ala11 in the wild-type (sheet) fold using the near-UV CD spectra of the same position-11 mutant in the Asn12 and Leu8 backgrounds as the helical and sheet basis spectra.





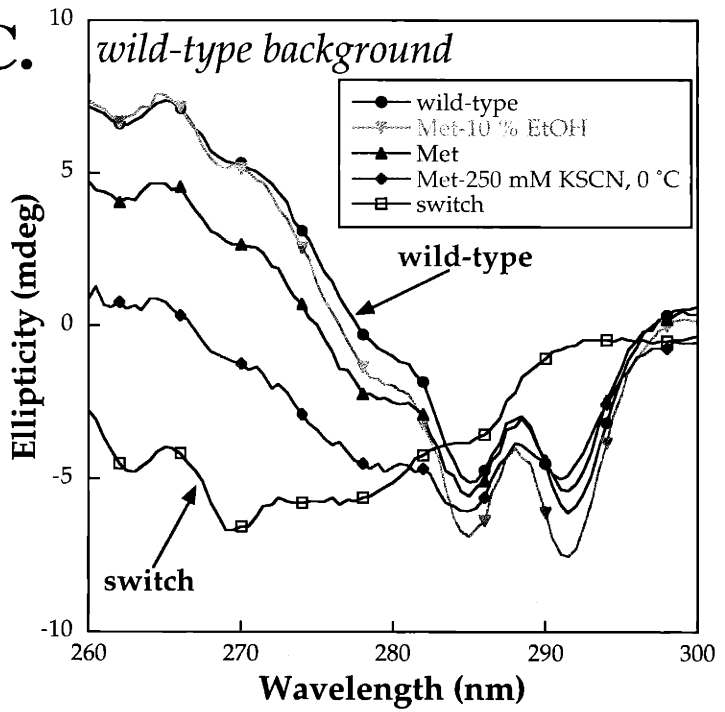


Figure 8. Near-UV CD spectra of Arc position-11 mutants under conditions which shift the sheet/helix equilibrium to favor helix or sheet. The near-UV CD spectra in storage buffer, storage buffer + ethanol, and storage buffer + KSCN at low temperature are shown for Gly11 (**A**), Ile11 (**B**), and Met11 (**C**). These spectra were taken at 100 μM protein concentration and 15 °C unless noted. For all experiments, Arc variants were in a buffer containing 50 mM Tris, 250 mM KCl, 0.2 mM EDTA, pH 7.5 plus the indicated conditions.

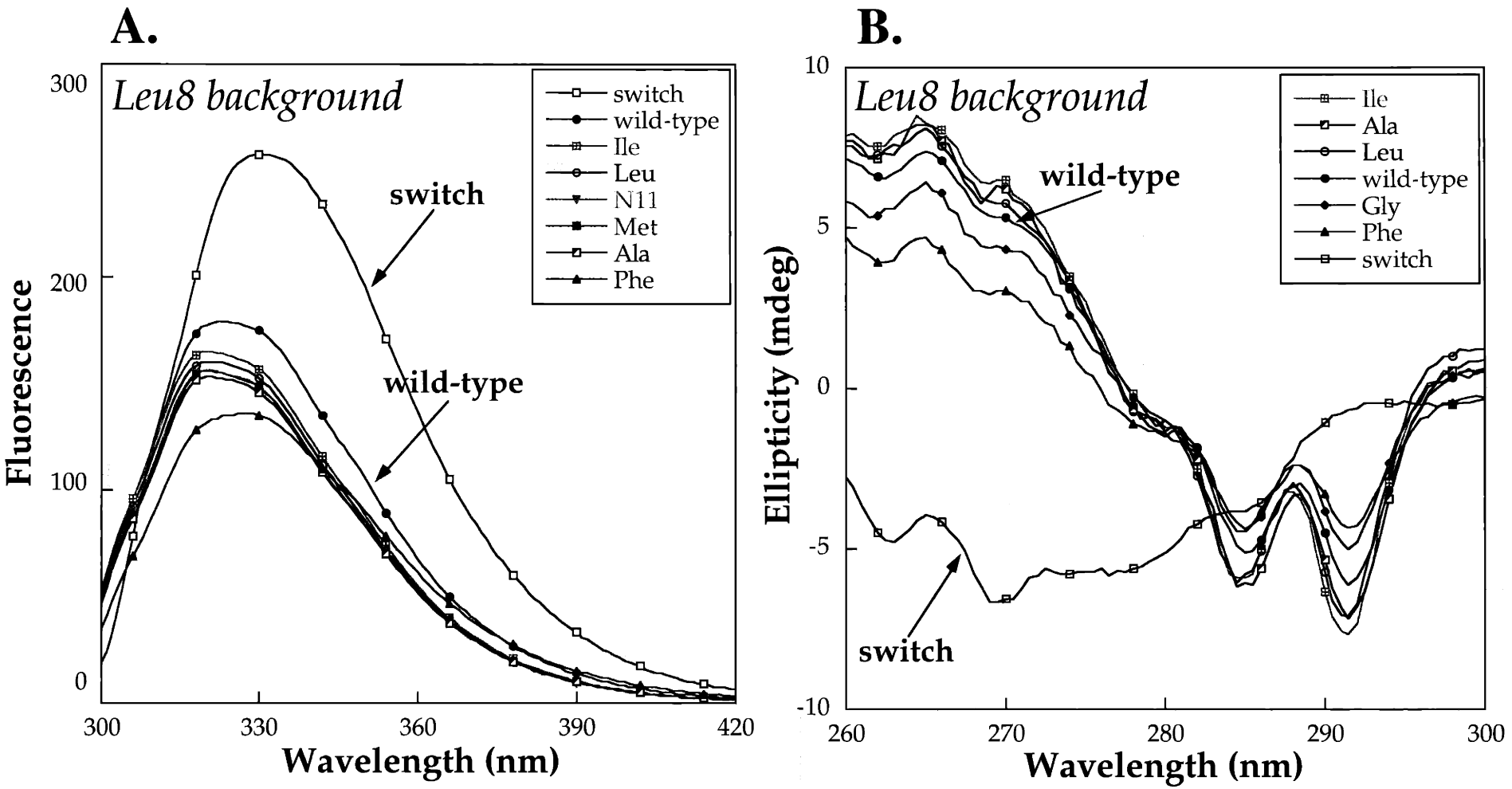
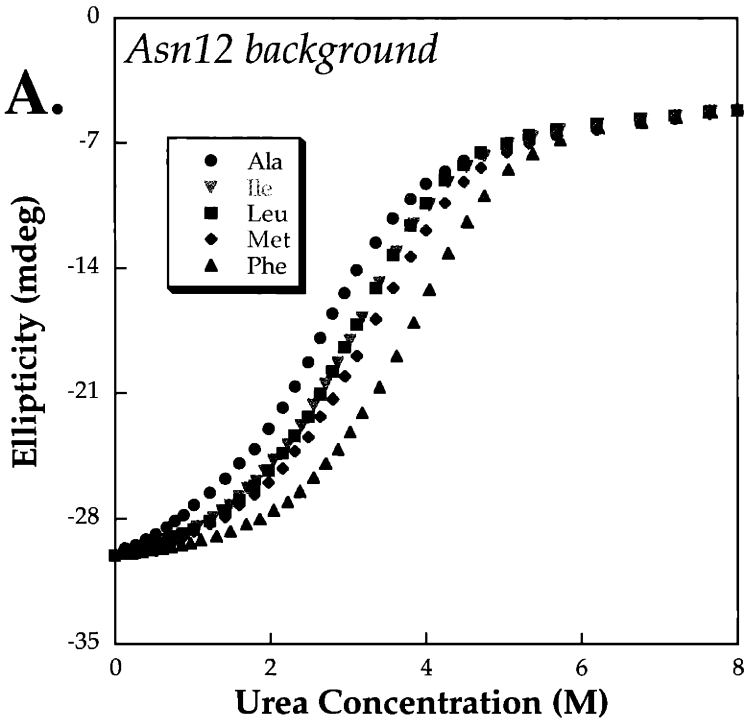


Figure 9. Fluorescence (**A**) and near-UV CD spectra (**B**) of switch Arc, wild-type Arc, and position-11 mutants in the Leu8 background. For all experiments, Arc variants were at 100 μM in a buffer containing 50 mM Tris, 250 mM KCl, 0.2 mM EDTA, pH 7.5. Fluorescence spectra were taken at 25 °C. Near-UV CD spectra were taken at 15 °C.



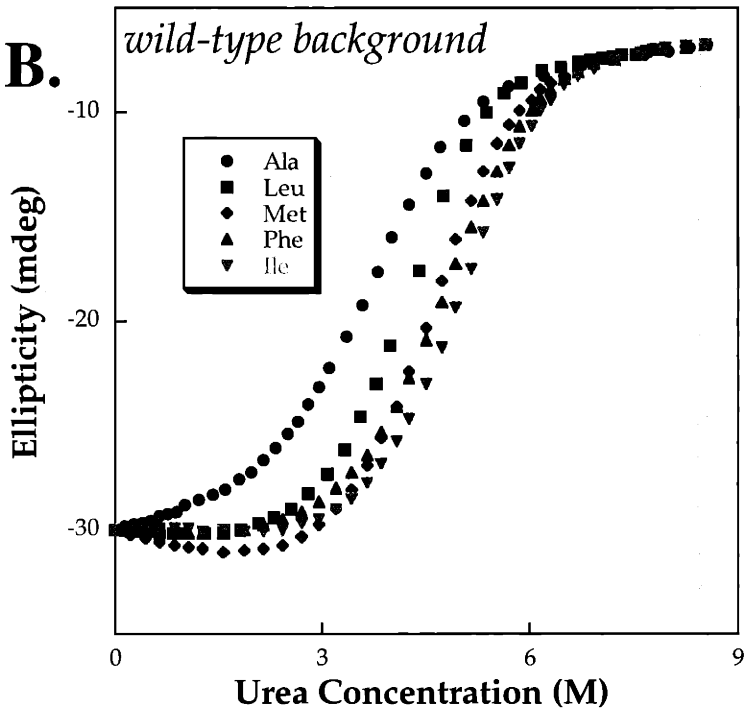
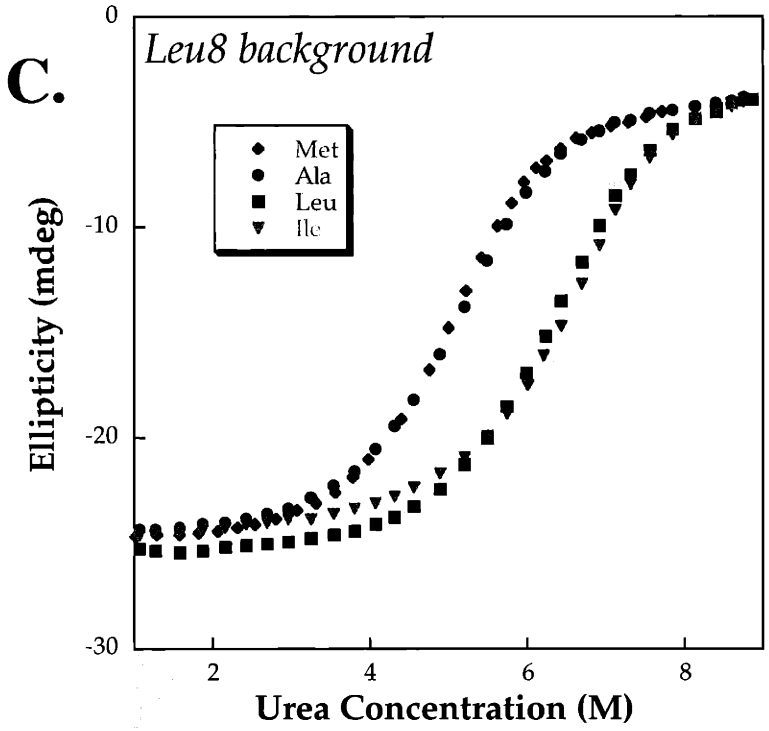


Figure 10. Urea denaturation curves for selected position-11 mutants (Ala, Ile, Phe, Leu, and Met) in the Asn12 (**A**), wild-type (**B**), and Leu8 (**C**) backgrounds. Data were taken at 5 μM protein concentration and 25 °C. All Arc variants were in a buffer containing 50 mM Tris, 250 mM KCl, 0.2 mM EDTA, pH 7.5.



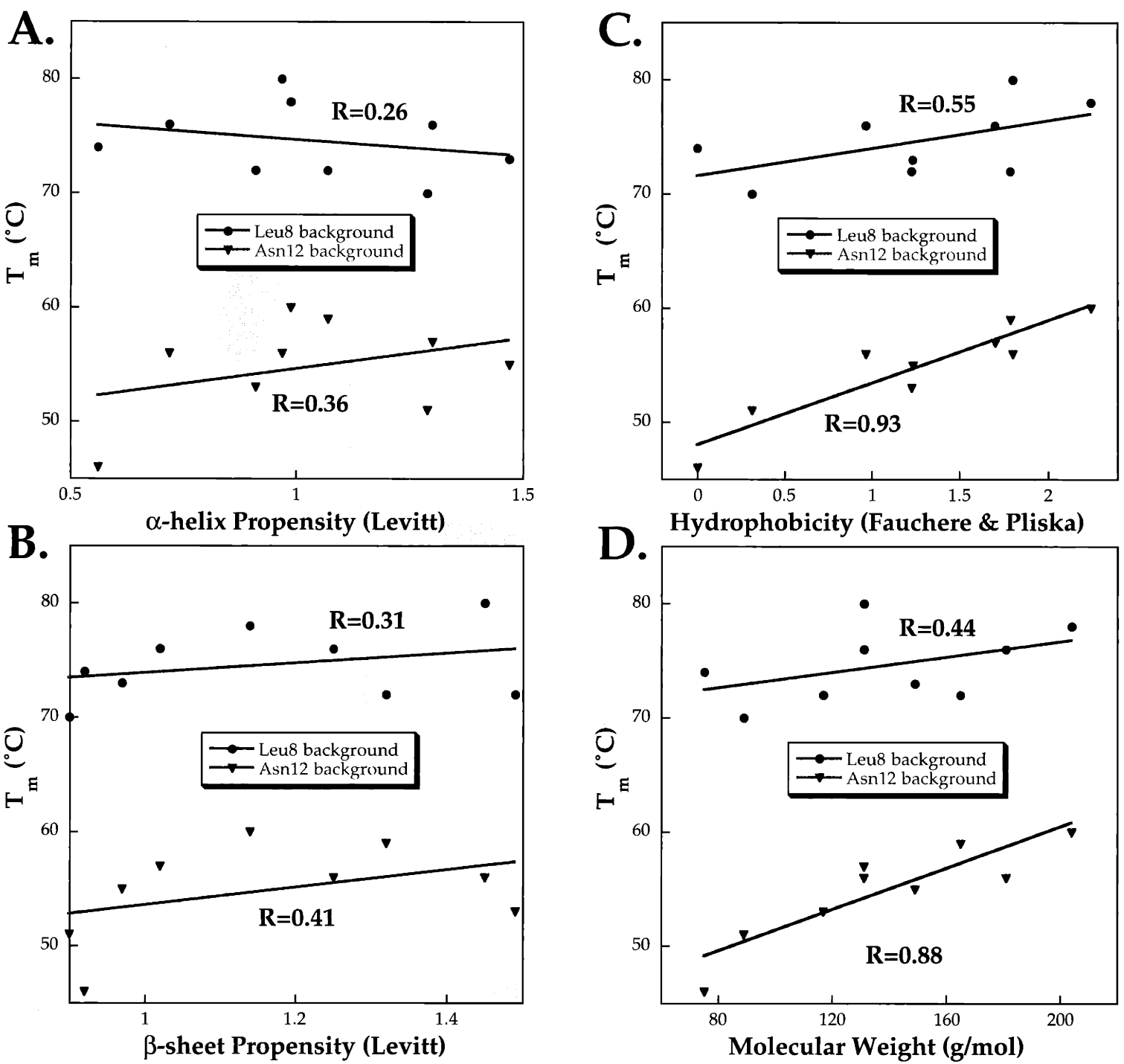


Figure 11. Correlation between the T_m's of position-11 mutants in the Asn12 and Leu8 backgrounds with α-helix propensity (**A**), β-sheet propensity (**B**), hydrophobicity (**C**), and molecular weight (**D**). (34, 35)

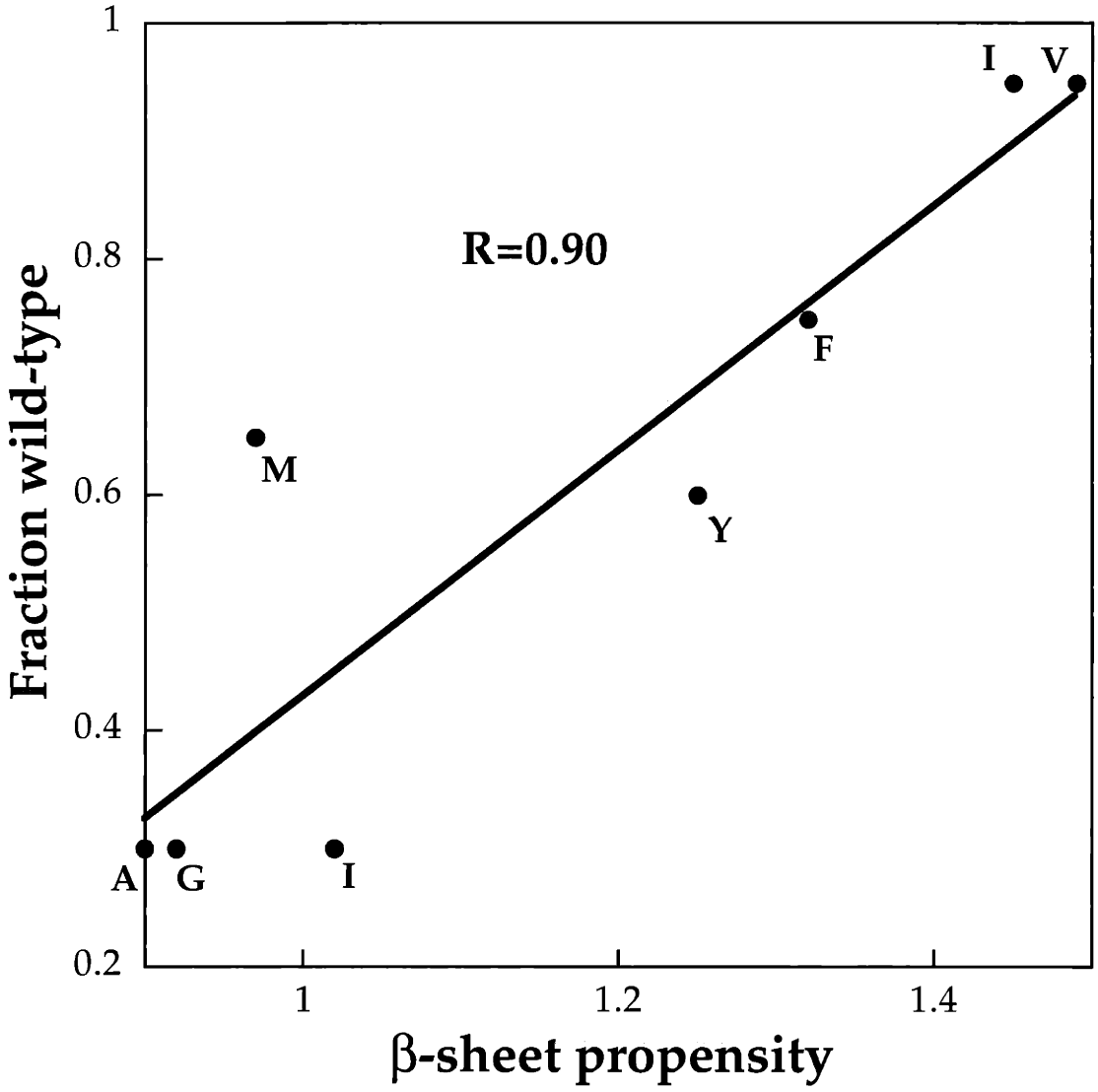


Figure 12. Correlation between the fraction of position-11 mutants (wild-type background) in the sheet fold with β-sheet propensity. (34)

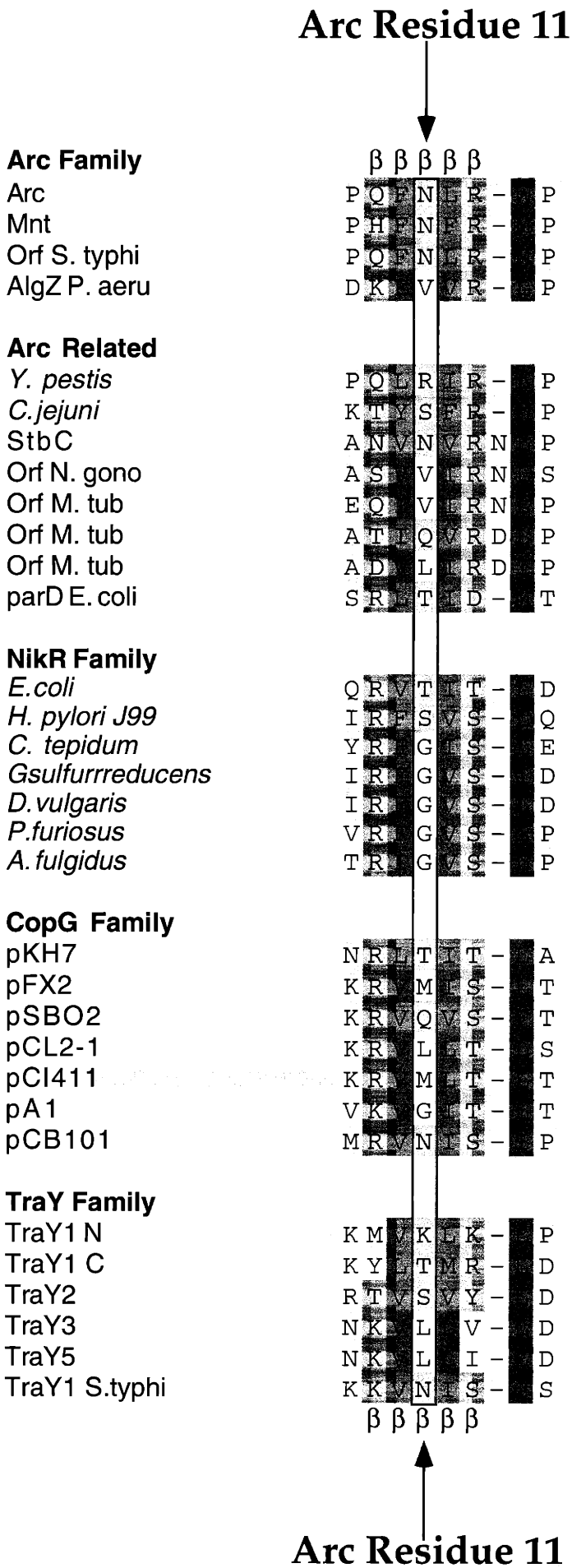


Figure 13. Alignment of the N-terminal region of homologous ribbon-helix-helix domains. (courtesy of P. Chivers)

Table 1. *Thermal stability of Arc position-11 mutants in three backgrounds.*

	background:		
residue 11:	**Asn12[a]**	**wt[a]**	**Leu8[a]**
Gly	46	56	70
Ala	51	63	72
Val	52	71	N.D.
Leu	57	67	76
Ile	56	71	80
Met	55	67	73
Phe	59	68	N.D.
Tyr	56	66	76

[a] thermal stability units are °C. The wild-type (N11, L12) T_m is 58°C. The T_m of Arc PL8 is 72°C. T_m's were determined at 10 uM protein concentration in a buffer containing 50 mM Tris, 250 mM KCl, 0.2 mM EDTA, pH 7.5.

Table 2. *Fraction of molecules in the wild-type conformation (storage buffer, 15°C).*

residue 11	fraction wild-type
Val	>0.95
Ile	>0.95
Phe	0.75
Met	0.65
Tyr	0.60
Ala	0.30
Gly	0.30
Leu	0.30

[a] The approximate fraction of each Arc position 11 mutant in the wild-type (sheet) conformation was determined by comparison with simulated near-UV CD spectra derived using near-UV spectra of the appropriate mutant in the Leu8 background and the Asn12 background as the sheet and helical basis spectra.

Table 3. *Free energy of unfolding/dissociation (kcal/mol) for Arc position-11 mutants in three backgrounds.*

	background:		
residue 11:	**Asn12[a, b]**	**wt[a, b]**	**Leu8[a, b]**
Ala	10.6	11.9	13.7
Leu	11.0	12.5	15.2
Ile	11.1	13.5	15.8
Met	11.3	12.9	13.6
Phe	12.0	13.2	N.D.

[a] m-value is fixed at -1.3. [b] units of free energy are kcal/mol. The ΔG of wild-type (N11, L12) Arc is 10.9 kcal/mol. The Arc PL8 ΔG is 13.6 kcal/mol. Free energy was determined by fitting chemical denaturation curves. Chemical denaturation was carried out at 5 uM protein concentration in a buffer containing 50 mM Tris, 250 mM KCl, 0.2 mM EDTA, pH 7.5 and urea from 0-9M at 25 °C.

Chapter 3

The Role of a N_{Cap} Residue in Arc's Stability and Activity

Introduction

It is important to understand how specific interactions and structural motifs contribute to the structure, stability, and activity of proteins. Helix-capping motifs, which prevent continuation of the helix and therefore help to establish a new trajectory for the polypeptide chain, can play significant roles in protein structure and stability (1-13). Such capping motifs may aid protein folding by reducing the conformational space a polypeptide side chain must sample to find the native fold of lowest free energy (14).

The N-capping box is an α-helix capping motif characterized by a pair of reciprocal side-chain to main-chain hydrogen bonds and a hydrophobic "staple motif" involving two side chains (14,15). At the N-terminus of an α-helix, the residue that makes an intrahelical hydrogen bond but does not have helical phi/psi angles is defined as the N_{cap} residue. Residues within the helix are numbered N1, N2, N3, etc. as one moves along the polypeptide chain away from the N_{cap} residue and toward the C-terminus of the helix. Residues before the helix are designated as N', N'', N''', etc. as one moves away from the N_{cap} residue toward the N-terminus of the protein (15,16). In the N-capping box, one hydrogen bond is typically present between the N_{cap} side chain and the N3 main-chain amide and another usually links the N_{cap} main-chain amide and the N3 side chain. In addition, the side chains of the N' and N4 residues of the N-capping box often make hydrophobic interactions with each other (14).

In general, mutation of the surface or solvent-exposed side chains of a protein has relatively modest effects on protein stability, whereas mutation of residues buried in the

hydrophobic core is more deleterious (17-31). There are interesting exceptions, however, to this rule. For example, in an alanine scan of the P22 Arc repressor (20), only one surface residue—Ser32, the N-cap residue of α-helix B (32,33)—was identified as playing a significant role in protein stability.

Mutating Ser32 to alanine was found to decrease the stability of the Arc dimer, both by slowing the rate of protein folding and dimerization and by increasing the rate of dimer dissociation and denaturation (20,34). In the crystal structure, the side-chain hydroxyl group of Ser32 forms hydrogen bonds with both the main-chain nitrogen and the side-chain hydroxyl group of Ser35 (the N3 residue) (Fig. 1). The Ser35 side chain is too short to make a reciprocal hydrogen bond with the Ser32 main-chain nitrogen, and the helix-B capping motif of Arc is therefore somewhat different than a canonical N-capping box. In addition, although the N' (Arg31) and N4 (Glu36) side chains of helix B interact, as expected for a canonical N-capping box, they form a salt bridge rather than a hydrophobic interaction (Fig. 1). Interestingly, the N' (Met31) and N_4 (Tyr36) side chains of helix B in the hyperstable MYL mutant of Arc do form a canonical hydrophobic interaction (35,36). I will refer to the capping motif at the N-terminus of Arc's helix B as a "near" N-capping box. Other proteins also use serine as the N_{cap}/N3 residues of N-capping boxes (15).

Arc functions as a repressor of gene expression by binding, as a dimer of dimers or tetramer, to a 21 base-pair operator DNA site (37-39). In the cocrystal complex, there are four Ser32 side chains and each hydroxyl group is within 3.5 to 4.5 Å of a number of

backbone phosphate groups (33) (Figure 2). As a result, the side-chain of Ser32 could play some role in determining the affinity of the repressor-operator interaction. This appears to be the case. Despite its markedly reduced stability, the Ala32 mutant had a longer half-life than the wild-type protein in operator-DNA binding assays (40). This result suggests either that the wild-type Ser32 side chain makes unfavorable DNA contacts that are relieved by the mutation or that the smaller mutant Ala32 side chain allows new and better DNA contacts.

To what extent is the evolutionary choice of Ser32 a compromise between adequate stability of the active protein dimer and adequate DNA-binding affinity? To address this question, I created a library of position-32 Arc variants and screened these mutants for repressor activity *in vivo*, DNA-binding activity *in vitro*, and protein stability. Arc mutants in which position 32 was occupied by residues commonly found at N_{cap} positions were generally reasonably stable, although only one mutant was more stable than wild-type Arc. Strong DNA binding, however, required a small side chain at position 32.

Materials and Methods

Mutagenesis, Selections, and Screens. A library containing all possible position-32 codons was constructed by cassette mutagenesis in the *arc* gene of pSA700 and transformed into *Escherichia coli* strain UA2F. The pSA700 plasmid also contained an Arc-repressible P_{ant} promoter fused to a dominant streptomycin sensitivity gene. Strain UA2F contained a recessive streptomycin resistance gene and a *λimm21* prophage with a

fusion of the P_{ant} promoter to the gene (*cat*) encoding chloramphenicol acetyl transferase. Active Arc variants repress transcription of the *cat* gene and the streptomycin sensitivity gene rendering host cells sensitive to chloramphenicol (75 μg/ml) and resistant to streptomycin (50 μg/ml). Inactive Arc variants, by contrast, result in cells that are chloramphenicol resistant and streptomycin sensitive (41,42). Arc mutants with approximately 5% of the wild-type activity are able to pass the streptomycin selection (42). Purification of mutant proteins and determination of thermal stabilities were performed as described in the previous chapter.

DNA Mobility Shift Assays. End-labeling of one strand of the O1 *arc* operator using γ-^{32}P-labeled ATP and T4 polynucleotide kinase was carried out as previously described (43). Equilibrium binding assays were performed and analyzed using a minor variation of established methods (38,43). To avoid inactivation, purified Arc variants were stored in lyophilized form and resuspended in storage buffer on the day of the assay. Proteins were then diluted in binding buffer (10 mM Tris (pH 7.5), 3 mM $MgCl_2$, 250 mM KCl, 0.1 mM EDTA, 0.1 mg/ml BSA, and 0.02% Nonidet NP40) and incubated with 10 pM ^{32}P-labeled O1 DNA for 3 to 4 hours at 25 °C. Glycerol was added to 5% of the total sample volume just before loading samples onto a 7% polyacrylamide gel (acrylamide:bisacrylamide ratio of 29:1; gel dimensions were 19 cm X 16 cm) in 0.5 X TBE buffer. A separate lane of tracking dye (0.02% xylene cyanol and 0.02% bromophenol blue) was loaded alongside protein/DNA samples (43). Gels were electrophoresed at 300 V for a minimum of 30 minutes prior to loading. After loading, gels were run at 300 V until the dye migrated into the gel at which point the voltage was

changed to 200 V and the gels were run until the bromophenol blue was within several inches of the bottom of the gel (approximately 45 minutes). Gels were dried and exposed using Molecular Dynamics phosphorimager screens. Band intensities were quantified using ImageQuant software.

Gel-shift data were fit to a model in which unfolded Arc monomers are in equilibrium with folded dimers and DNA-bound tetramers.

$$4U \overset{K_u}{\Leftrightarrow} 2A_2 + O \overset{K_d}{\Leftrightarrow} A_4O$$

The equilibrium constant for dimer dissociation and denaturation is $K_u = [U]^2/[A_2]$, and the equilibrium constant for dissociation of the DNA-bound tetramer is $K_d = [A_2]^2[O]/[A_4O]$. The equilibrium constant for the overall reactions is $K_{obs} = K_u^2 K_d$, and the fraction of bound operator is $\theta = [A_4O]/([O] + [A_4O]) = 1/(1 + K_{obs}/[U]^4)$. Gel-shift data were fit using the nonlinear least-squares subroutine implemented in the program Kaleidagraph running on a Macintosh computer.

Results

Effects of position 32 replacement on Arc activity in vivo. Using cassette mutagenesis, I constructed a library in which all possible amino-acid substitutions were permitted at position 32 of Arc repressor. An activity assay in which repressor binding to its operator results in antibiotic resistance or sensitivity *in vivo* was then used to test individual clones (Table 1). In all, 14 different substitutions at position 32 were characterized. Only two mutants (Ala32 and Cys32) had intracellular activities similar to wild-type Arc (Ser32) by the criteria that host cells were resistant to streptomycin and sensitive to

chloramphenicol. Two additional mutants (Pro32 and Thr32) had partial activity and rendered the UA2F/pSA700 host partially resistant to streptomycin and to chloramphenicol. Mutants containing Asp32, Asn32, Glu32, Gln32, Gly32, His32, Met32, Val32 and Leu32 were inactive in the cell; their host cells were killed by streptomycin and were resistant to chloramphenicol.

Repressor activity in the cell depends on the expression level of the protein, the fraction of expressed protein that forms native Arc dimers, and the intrinsic operator-binding activity of the dimer. Mutants in which the stability of the Arc dimer is reduced are degraded more rapidly in the cell and thus have reduced expression levels (42,44,45). Mutations that decrease Arc stability also increase the fraction of denatured monomers relative to native dimers, which also decreases repressor activity. Thus, the inactivity *in vivo* of the Asp32, Asn32, Glu32, Gln32, Gly32, His32, Met32, Val32 and Leu32 variants of Arc could result from alterations in either stability or DNA affinity.

Thermal stability of purified variants. I purified each mutant protein and used far-UV circular dichroism (CD) to determine the fraction of native and denatured protein at different temperatures. At 15 °C, each of the mutants, like wild-type Arc, had a CD spectrum expected for a protein with roughly 60% α-helical structure (see Fig. 3A for selected spectra). Moreover, each mutant showed a cooperative thermal unfolding transition (Fig. 3B). Only one mutant, Asp32, was more thermally stable than wild-type Arc; it's T_m was increased by 7 °C (Fig. 3B; Table 2). The remaining mutants had T_m's reduced from 4 to 29 °C relative to wild type (Fig. 3B; Table 2). Clearly, the chemical

identity of residue 32 plays a large role in determining the intrinsic stability of the Arc dimer.

DNA-binding affinities of purified mutants. I assayed the operator-DNA binding of each of the purified position-32 variants using a gel mobility shift assay. Representative experiments and binding curves are shown in Fig. 4 for the Asn32, Cys32, and Thr32 mutants. Table 3 lists the protein concentrations required to observe half-maximal binding for each mutant. These values range from 0.6 nM for wild-type Arc (Ser32) to 33 μM for the Asp32 variant. Because the Asp32 variant is more stable than wild-type Arc, its dramatic reduction in operator binding affinity must result directly from unfavorable contacts with operator DNA. Pro32, Cys32, Asn32, and Ala32 were the most active variants, with overall DNA-binding activities within 10-fold of the wild-type proteins.

Discussion

The results presented here show that the N_{cap} residue of helix B in Arc repressor plays an important role in determining the stability of this protein to thermal denaturation and also in establishing its operator-binding affinity. Roughly 70% of the surface of the wild-type Ser32 side chain is exposed to solvent and mutant side chains at this position would also be highly solvent accessible. Nevertheless, in a set of 14 position-32 variants, the T_m's ranged from a high of 71 °C for the Asp32 mutant to a low of 35 °C for the Leu32 mutant. As a result, N_{cap} positions seem to be an exception to the general rule that surface residues do not play major roles in determining protein stability.

The great majority of the most stabilizing residues at position 32 in Arc are also commonly found at the N_{cap} positions in α-helices of other proteins and *vice versa*. The T_m's of the position-32 variants in Arc correlate reasonably well (R = 0.73) with the frequencies of N_{cap} occurrence of the side chain in a library of 2101 α-helices in high-resolution protein structures (46). In Arc, Asp32, Ser32, Thr32, and Asn32 represent four of the five most stable variants. The side chains of each of these residues can form hydrogen bonds with the main-chain nitrogen of the N3 residue (Ser35), a defining feature of the N-capping box motif. Among these four residues, Asp32 was the most stabilizing in Arc, probably because its negative charge interacts favorably with the partial positive charge at the N-terminus of the α-helix.

The Cys32 mutant had the third highest thermal stability in the library of position-32 Arc variants. This does not seem surprising given that the cysteine side chain is sterically most similar to the wild-type serine side chain, is capable of forming weak hydrogen bonds (47,48), and in its deprotonated state could also interact favorably with the positive end of the helical dipole. Remarkably, however, cysteine is the rarest N_{cap} residue in natural proteins, comprising less than 1% of all of these positions (46,49). Relative to other protein helices, there is nothing obvious about the structure surrounding the N_{cap} position of helix-B of Arc that would suggest that this is a unique or highly favorably environment for cysteine. If this is true, however, then cysteine should also be a good N_{cap} residue in other proteins. Why then is cysteine used so rarely in this capacity? One intriguing possibility is that the interaction with the helical dipole in an

N_{cap} context reduces the pK_a of the cysteine side chain, resulting in a higher population of the deprotonated species. Because this species is the reactive form in nucleophilic displacements, solvent-exposed cysteine side chain at N_{cap} positions might be highly reactive and potentially prone to oxidation and other types of chemical modification. Interestingly, when Cys32 is removed from the comparison of T_m's in the Arc variants with N_{cap} frequencies, the correlation coefficient improves from 0.73 to 0.88 (Fig. 5).

The identity of the side chain at position 32 in Arc also plays a critical role in function by influencing operator-binding affinity. In the wild-type protein-DNA complex, Ser32 is close to the DNA backbone and the side-chain hydroxyl group is relatively solvent inaccessible. Although the wild-type Ser32 side chain does not make any direct hydrogen bonds with the operator, mutant side chains at position 32 could destabilize the complex via steric or electrostatic clashes with the DNA. For example, modeling Thr32 into the protein-DNA complex without allowing structural relaxation results in steric clashes (2.4 Å) between the γ-CH_3 of the mutant side chain and the phosphate backbone. Indeed, even though the Thr32 Arc variant is quite thermally stable, this mutant is only partially active *in vivo* and displays a significant reduction in operator binding *in vitro*. Not surprisingly, most large and/or negatively charged side chains at position 32 (His, Glu, Gln, Met, Leu, Asp) resulted in proteins that were inactive in the cell and displayed dramatic reductions in operator binding *in vitro*. The Pro32, Cys32, Asn32, and Ala32 variants had activities within 10-fold of the wild-type Ser32 protein *in vitro*, and, except for Asn32, showed some activity *in vivo*. Thus, relatively small and uncharged side chains appear to be accommodated in the protein-

DNA complex better than larger ones. The Pro32 mutant has the same activity as wild-type Arc *in vitro* despite its 15 °C reduction in thermal stability. In fact, after correcting for the difference in stability, which should reduce the apparent activity roughly 10-fold, the Pro32 variant is slightly more active than wild type. This may occur because the Pro ring is able to make additional packing interactions with the DNA backbone or results in a conformational change which improves other contacts.

The choice of side chain for position 32 in Arc repressor clearly represents an evolutionary compromise between protein stability and DNA-binding activity. The wild-type residue, Ser32, does not result in the most stable protein nor in the highest intrinsic operator affinity but it is ranked near the top of both categories. Variants with Asp32 would be more stable but inactive and variants with Pro32 or Ala32 would probably be too unstable to permit strong DNA binding in the cell. Arc is a member of the ribbon-helix-helix family of transcription factors, and it is instructive to ask how other family members choose the N_{cap} residue for helix B. In a set of 58 paralogs—including the MetJ, CopG, and TraY subfamilies—serine (55%), threonine (33%), and proline (7%) accounted for 95% of all of the helix-B N_{cap} residues. These results are consistent with those presented here, in that each of these side chains shows some activity in the Arc background. In initial sequence comparisons, I used the published alignment of MetJ with other ribbon-helix-helix proteins, which aligns Ala50 of MetJ with Ser32 of Arc as the helix-B N_{cap} (Fig. 6). Inspection of the MetJ crystal structure (50), however, clearly showed that Thr51 is the N_{cap} residue for helix B of this protein, and the alignments were changed to reflect this fact (Fig. 6).

Somewhat different results were observed in the NikR ribbon-helix-helix subfamily (Fig. 6), where glutamine (26%) and asparagine (26%) were most common and glutamic acid (5%) and methionine (5%) were also observed among 19 subfamily members. With the exception of asparagine, these residues result in unstable, inactive Arc proteins. Unfortunately, structures for NikR and its operator complex are not presently available. Perhaps these structures would reveal why N_{cap} residues that would be highly destabilizing and functionally deleterious in Arc are allowed in NikR. A more likely possibility, however, is that the current sequence alignment between NikR family members and other ribbon-helix-helix proteins is offset by one residue at the beginning of helix B, as was the case for MetJ. In this case, serine (38%) and asparagine (50%) would occupy most of the helix-B N_{cap} positions in NikR (Fig. 6). Hence, knowledge of the structural and functional effects of many amino acid substitutions at an important sequence position in one family member, such as position 32 in Arc, may help to constrain sequence alignments with more distantly related proteins.

References

1. Iovino, M., Falconi, M., Petruzzelli, R., and Desideri, A. (2001) *J Biomol Struct Dyn* **19**(2), 237-46.

2. elMasry, N. F., and Fersht, A. R. (1994) *Protein Eng* **7**(6), 777-82.

3. Lanigan, M. D., Tudor, J. E., Pennington, M. W., and Norton, R. S. (2001) *Biopolymers* **58**(4), 422-36.

4. Laity, J. H., Chung, J., Dyson, H. J., and Wright, P. E. (2000) *Biochemistry* **39**(18), 5341-8.

5. Lu, M., Shu, W., Ji, H., Spek, E., Wang, L., and Kallenbach, N. R. (1999) *J Mol Biol* **288**(4), 743-52.

6. Motoshima, H., Mine, S., Masumoto, K., Abe, Y., Iwashita, H., Hashimoto, Y., Chijiiwa, Y., Ueda, T., and Imoto, T. (1997) *J Biochem (Tokyo)* **121**(6), 1076-81.

7. Parker, M. H., and Hefford, M. A. (1998) *Biotechnol Appl Biochem* **28**(Pt 1), 69-76.

8. Serrano, L., and Fersht, A. R. (1989) *Nature* **342**(6247), 296-9.

9. Wilcock, D., Pisabarro, M. T., Lopez-Hernandez, E., Serrano, L., and Coll, M. (1998) *Acta Crystallogr D Biol Crystallogr* **54**(Pt 3), 378-85.

10. Zhukovsky, E. A., Mulkerrin, M. G., and Presta, L. G. (1994) *Biochemistry* **33**(33), 9856-64.

11. Dragani, B., Stenberg, G., Melino, S., Petruzzelli, R., Mannervik, B., and Aceto, A. (1997) *J Biol Chem* **272**(41), 25518-23.

12. Cocco, R., Stenberg, G., Dragani, B., Rossi Principe, D., Paludi, D., Mannervik, B., and Aceto, A. (2001) *J Biol Chem* **276**(34), 32177-83.

13. Rossjohn, J., McKinstry, W. J., Oakley, A. J., Parker, M. W., Stenberg, G., Mannervik, B., Dragani, B., Cocco, R., and Aceto, A. (2000) *J Mol Biol* **302**(2), 295-302.

14. Munoz, V., Blanco, F. J., and Serrano, L. (1995) *Nat Struct Biol* **2**(5), 380-5.

15. Harper, E. T., and Rose, G. D. (1993) *Biochemistry* **32**(30), 7605-9.

16. Presta, L. G., and Rose, G. D. (1988) *Science* **240**(4859), 1632-41.

17. Herrmann, L., Bowler, B. E., Dong, A., and Caughey, W. S. (1995) *Biochemistry* **34**(9), 3040-7.

18. Pakula, A. A., and Sauer, R. T. (1990) *Nature* **344**(6264), 363-4.

19. Reidhaar-Olson, J. F., and Sauer, R. T. (1990) *Proteins* **7**(4), 306-16

20. Milla, M. E., Brown, B. M., and Sauer, R. T. (1994) *Nat Struct Biol* **1**(8), 518-23.

21. Michael, S. F., Kilfoil, V. J., Schmidt, M. H., Amann, B. T., and Berg, J. M. (1992) *Proc Natl Acad Sci U S A* **89**(11), 4796-800.

22. Shang, Z., Isaac, V. E., Li, H., Patel, L., Catron, K. M., Curran, T., Montelione, G. T., and Abate, C. (1994) *Proc Natl Acad Sci U S A* **91**(18), 8373-7.

23. Shortle, D., Stites, W. E., and Meeker, A. K. (1990) *Biochemistry* **29**(35), 8033-41.

24. Serrano, L., Kellis, J. T., Jr., Cann, P., Matouschek, A., and Fersht, A. R. (1992) *J Mol Biol* **224**(3), 783-804.

25. Eriksson, A. E., Baase, W. A., and Matthews, B. W. (1993) *J Mol Biol* **229**(3), 747-69.

26. Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P., and Matthews, B. W. (1992) *Science* **255**(5041), 178-83.

27. Lim, W. A., and Sauer, R. T. (1989) *Nature* **339**(6219), 31-6.

28. Lim, W. A., Farruggio, D. C., and Sauer, R. T. (1992) *Biochemistry* **31**(17), 4324-33.

29. Lim, W. A., and Sauer, R. T. (1991) *J Mol Biol* **219**(2), 359-76.

30. Lim, W. A., Hodel, A., Sauer, R. T., and Richards, F. M. (1994) *Proc Natl Acad Sci U S A* **91**(1), 423-7.

31. Cordes, M. H., and Sauer, R. T. (1999) *Protein Sci* **8**(2), 318-25.

32. Breg, J. N., van Opheusden, J. H., Burgering, M. J., Boelens, R., and Kaptein, R. (1990) *Nature* **346**(6284), 586-9.

33. Raumann, B. E., Rould, M. A., Pabo, C. O., and Sauer, R. T. (1994) *Nature* **367**(6465), 754-7.

34. Milla, M. E., Brown, B. M., Waldburger, C. D., and Sauer, R. T. (1995) *Biochemistry* **34**(42), 13914-9.

35. Waldburger, C. D., Schildbach, J. F., and Sauer, R. T. (1995) *Nat Struct Biol* **2**(2), 122-8.

36. Nooren, I. M., Rietveld, A. W., Melacini, G., Sauer, R. T., Kaptein, R., and Boelens, R. (1999) *Biochemistry* **38**(19), 6035-42.

37. Vershon, A. K., Youderian, P., Susskind, M. M., and Sauer, R. T. (1985) *J Biol Chem* **260**(22), 12124-9.

38. Brown, B. M., Bowie, J. U., and Sauer, R. T. (1990) *Biochemistry* **29**(51), 11189-95.

39. Vershon, A. K., Liao, S. M., McClure, W. R., and Sauer, R. T. (1987) *J Mol Biol* **195**(2), 323-31.

40. Brown, B. M., Milla, M. E., Smith, T. L., and Sauer, R. T. (1994) *Nat Struct Biol* **1**(3), 164-8.

41. Bowie, J. U., and Sauer, R. T. (1989) *J Biol Chem* **264**(13), 7596-602.

42. Bowie, J. U., and Sauer, R. T. (1989) *Proc Natl Acad Sci U S A* **86**(7), 2152-6.

43. Brown, B. M., and Sauer, R. T. (1993) *Biochemistry* **32**(5), 1354-63.

44. Milla, M. E., Brown, B. M., and Sauer, R. T. (1993) *Protein Sci* **2**(12), 2198-205.

45. Vershon, A. K., Bowie, J. U., Karplus, T. M., and Sauer, R. T. (1986) *Proteins* **1**(4), 302-11.

46. Penel, S., Hughes, E., and Doig, A. J. (1999) *J Mol Biol* **287**(1), 127-43.

47. Gregoret, L. M., Rader, S. D., Fletterick, R. J., and Cohen, F. E. (1991) *Proteins* **9**(2), 99-107.

48. Luscombe, N. M., Laskowski, R. A., and Thornton, J. M. (2001) *Nucleic Acids Res* **29**(13), 2860-74.

49. Penel, S., Morrison, R. G., Mortishire-Smith, R. J., and Doig, A. J. (1999) *J Mol Biol* **293**(5), 1211-9.

50. Rafferty, J. B., Somers, W. S., Saint-Girons, I., and Phillips, S. E. (1989) *Nature* **341**(6244), 705-10.

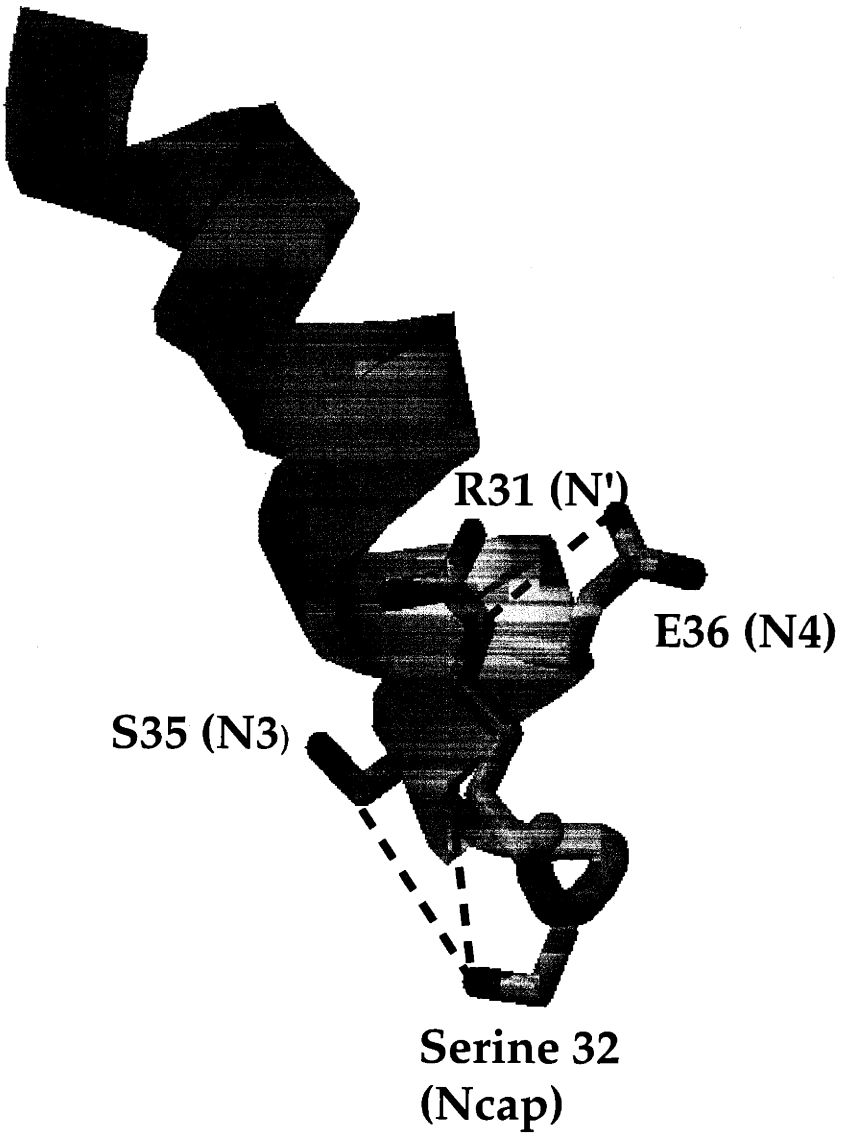


Figure 1. The 'near' N-capping box of Arc's helix B . Residues 31 to 36 of Arc participate in a 'near' N-capping box in which the serine 32 side-chain hydrogen bonds to the backbone amide of serine 35 and a salt bridge exists between arginine 31 (N') and glutamate 36 (N4). In addition, the serine 32 side-chain hydrogen bonds to the serine 35 side-chain. (33)

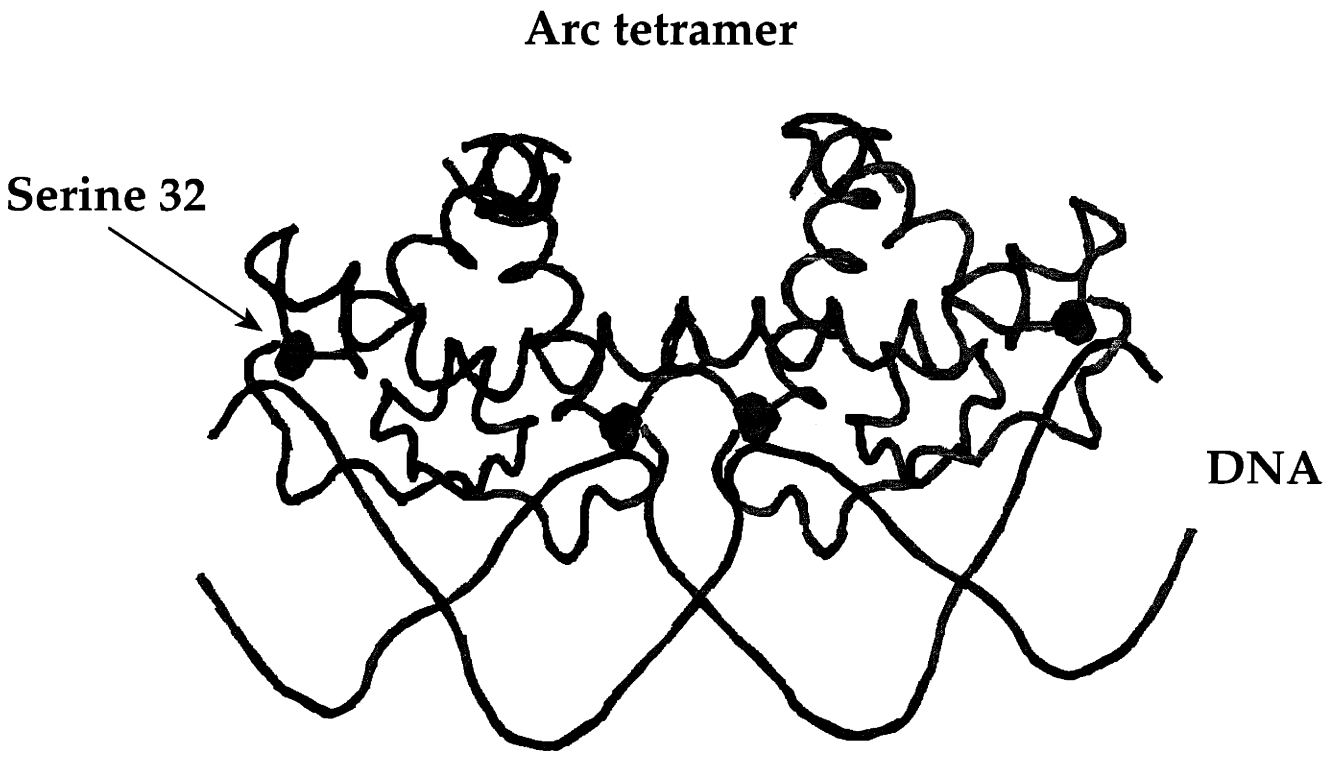


Figure 2. The Arc tetramer is pictured bound to operator DNA. Ser32 from each monomer is shown in CPK representation. The serine side-chain oxygen (red) is close to the DNA backbone. (33)

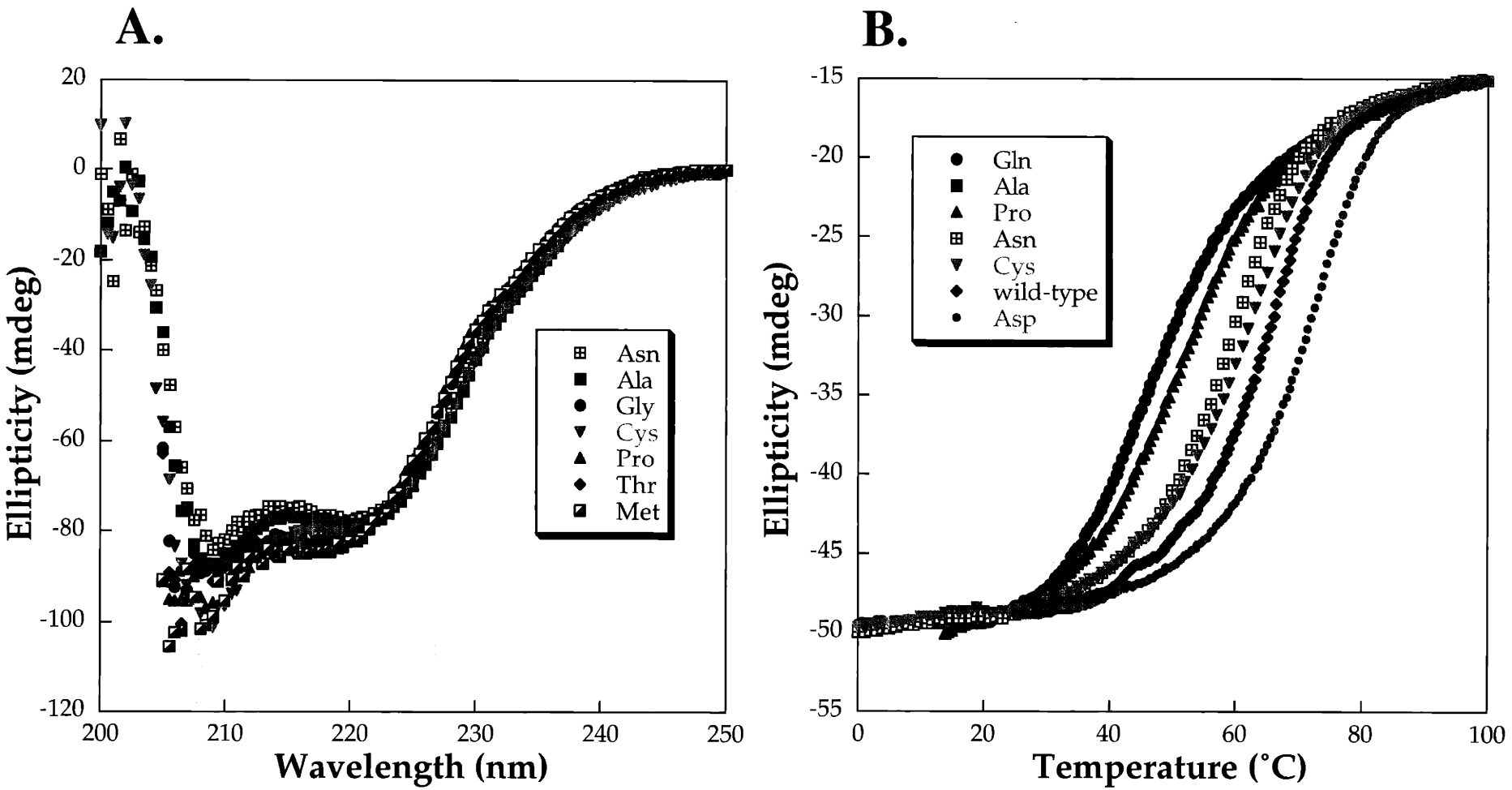


Figure 3. Far-UV CD spectra (**A**) and thermal denaturation curves (**B**) of position-32 mutants. Far-UV CD spectra were taken at 10 μM protein concentration and 15 °C. Thermal melts were taken at 50 μM protein concentration. For all experiments, Arc variants were in a buffer containing 50 mM Tris, 250 mM KCl, 0.2 mM EDTA, pH 7.5.

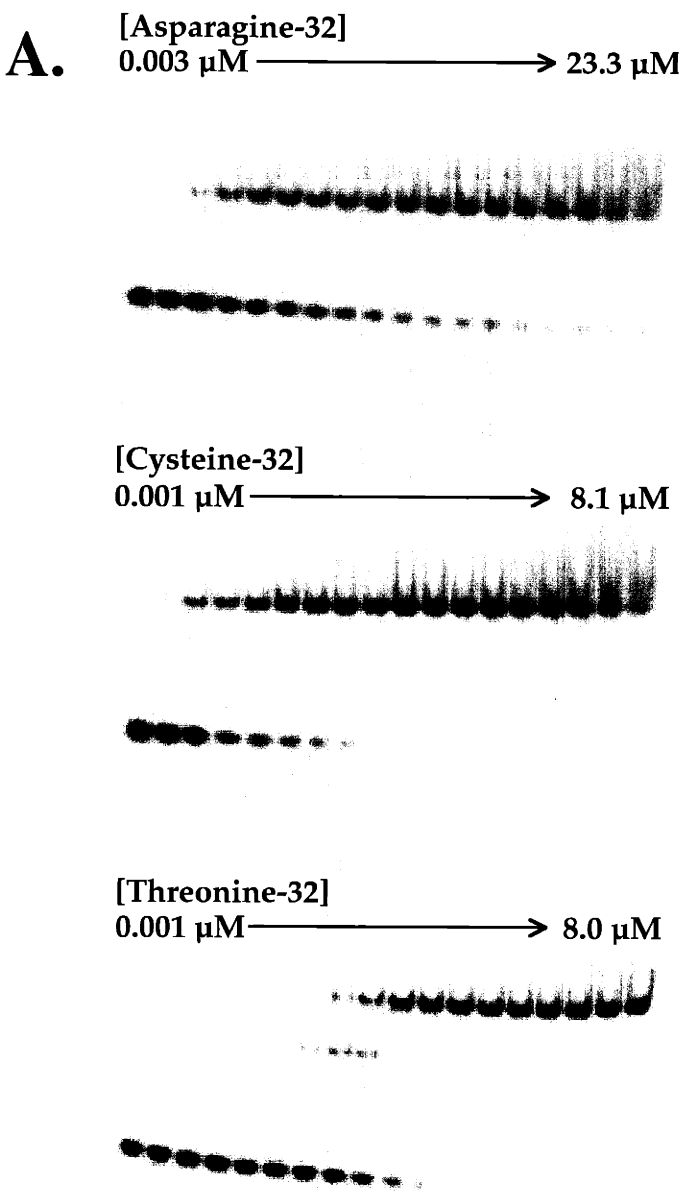
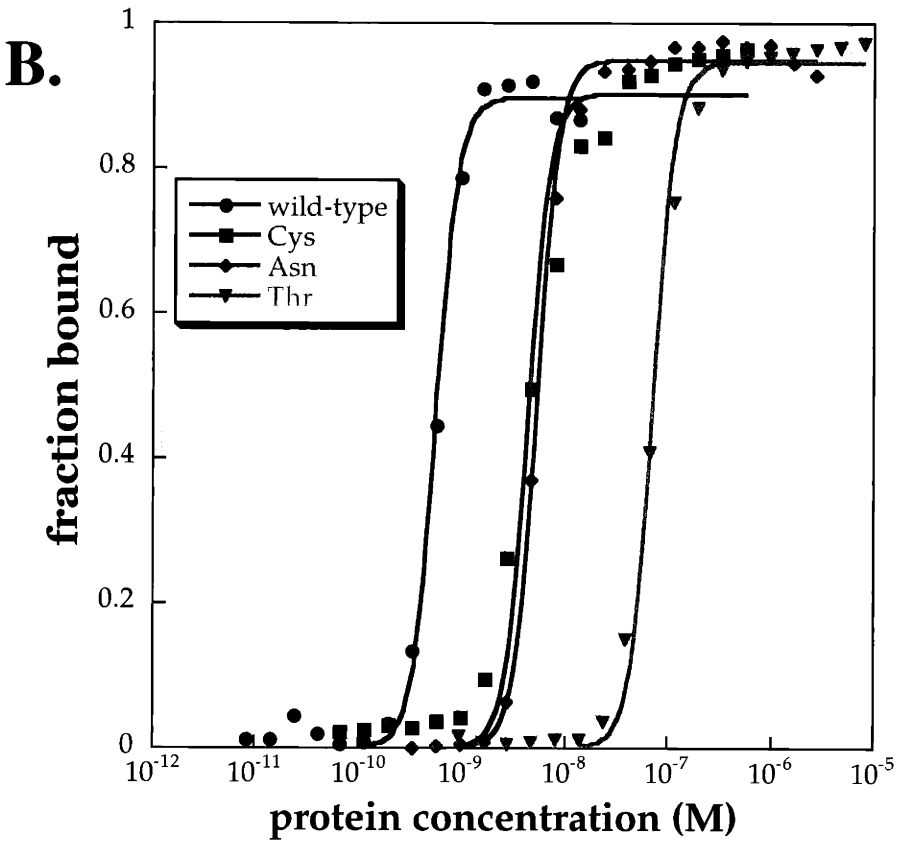


Figure 4. DNA mobility shift assays of Arc position 32 mutants. Gel shift autoradiograms are shown in (**A**) while the fraction of DNA bound is plotted against protein concentration in (**B**). ^{32}P-labelled oligonucleotide containing the Arc operator (O1) was used for these experiments. The protein concentration was increased in 1.7-fold increments starting and ending with the protein concentrations shown. Protein and DNA were incubated at 25 °C in a buffer containing 10 mM Tris pH 7.5, 3 mM $MgCl_2$, 250 mM KCl, 0.1 mM EDTA, 0.1 mg/ml BSA, and 0.02% Nonidet NP-40.



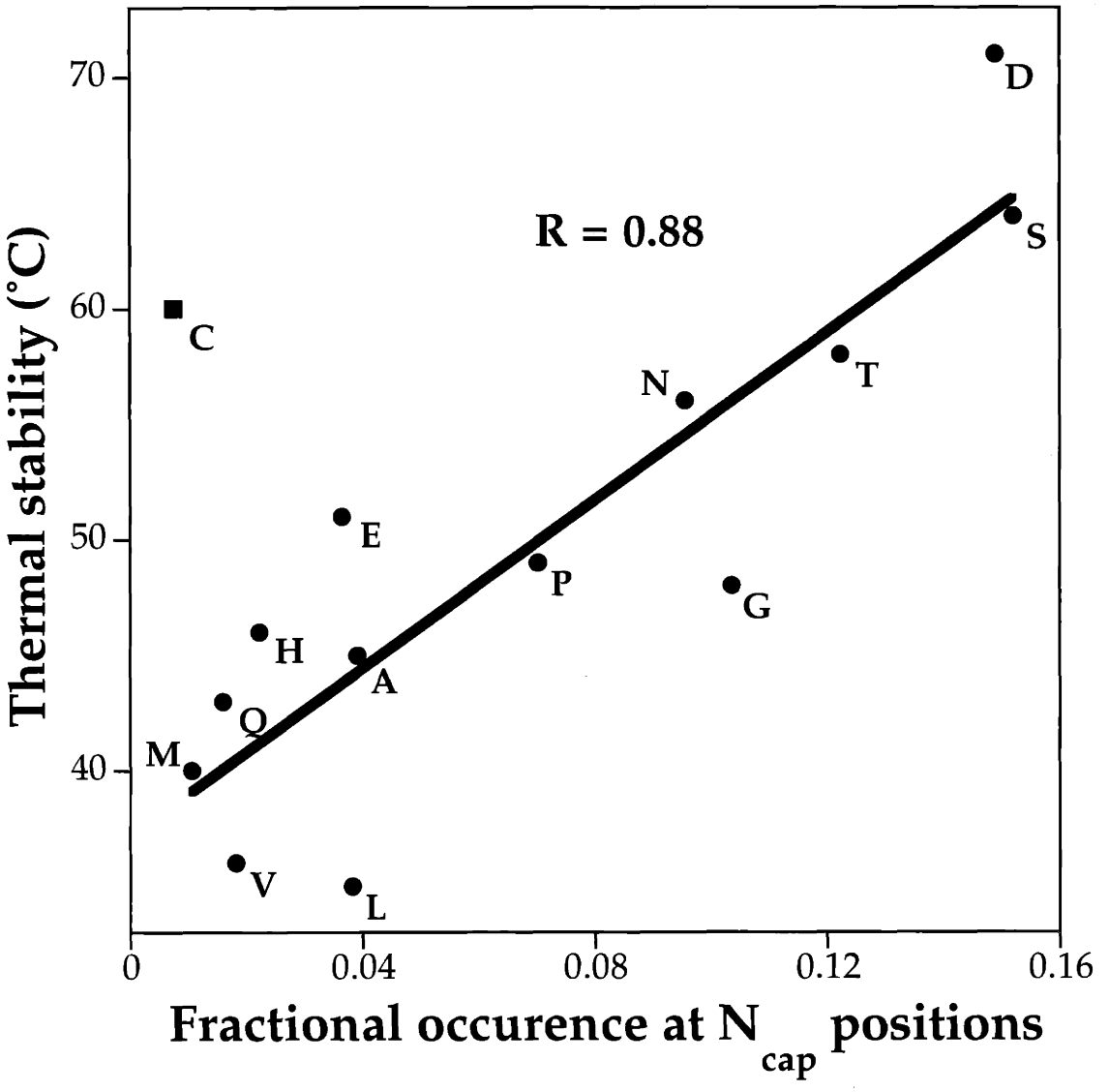


Figure 5. The correlation between the thermal stability of Arc position-32 mutants and their fractional occurrence at N_{cap} positions as determined using a library of 2101 α-helices in high-resolution protein structures. The Cys data point (red) was not used for this fit. (46)

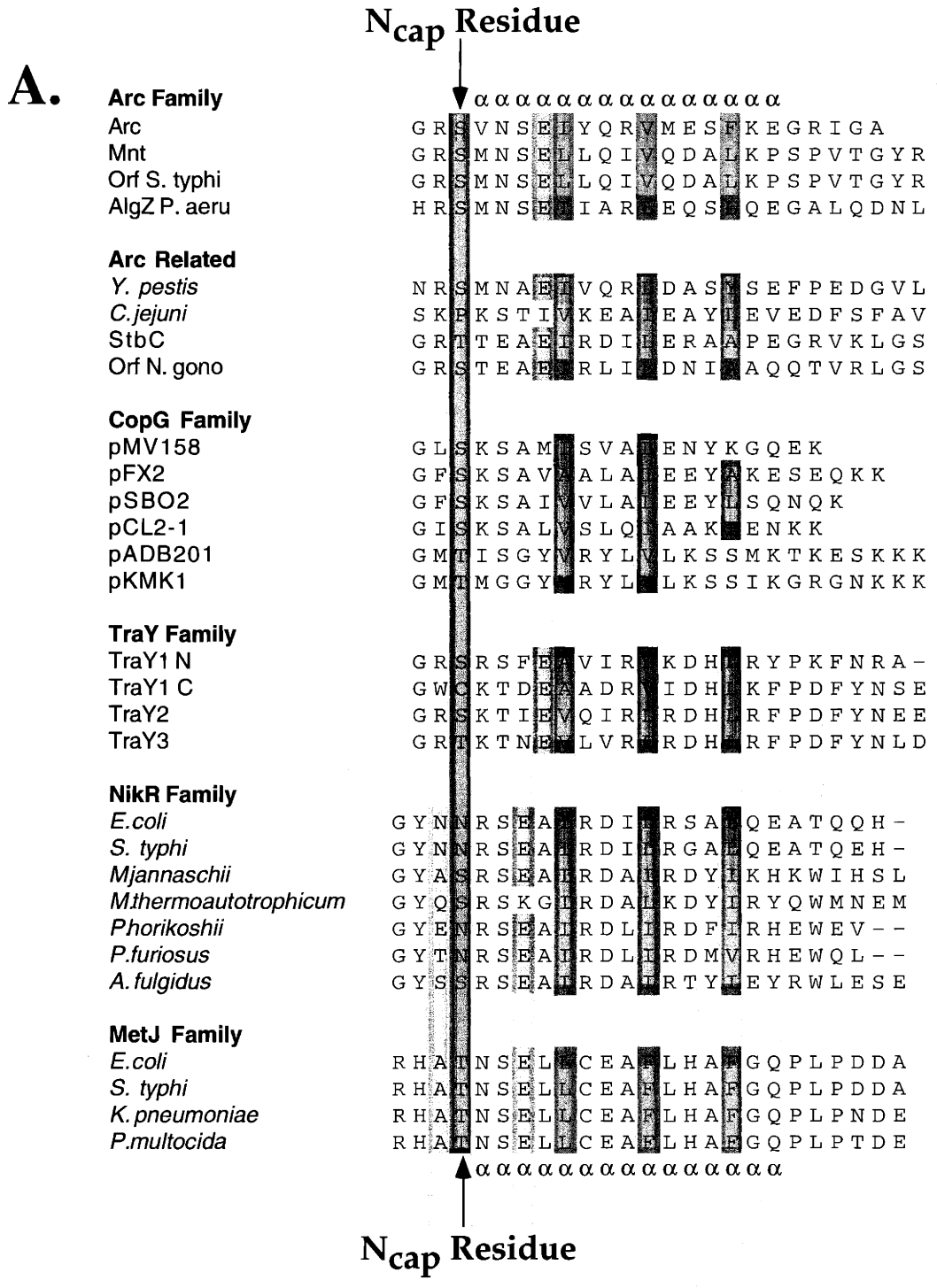


Figure 6. Alignment of the C-terminal region of homologous ribbon-helix-helix domains (**A,** courtesy of P. Chivers) and common N_{cap}-N3 partners (**B,** 15). The N_{cap} residue from each sequence (**A**) is labelled and boxed in blue. The N_{cap} residue according to previous alignment is boxed in light green.

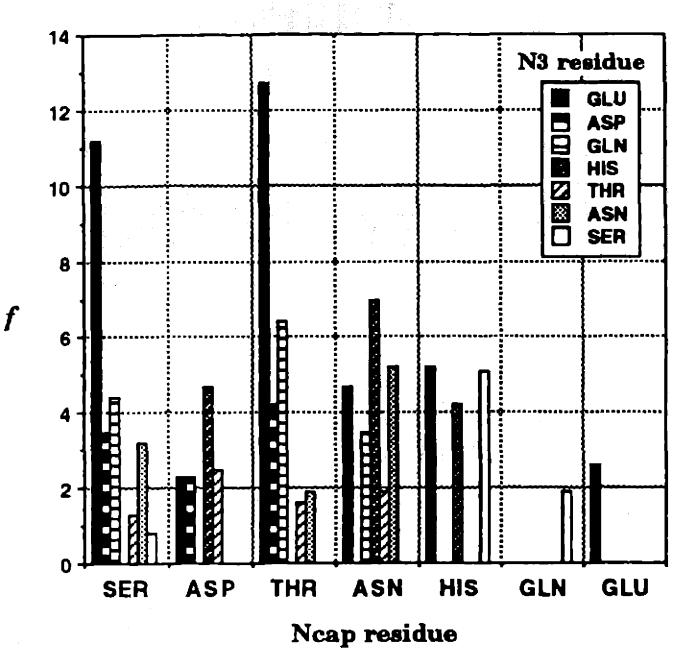


Table 1. *In vivo activity of Arc position-32 mutants.*

residue 32	*in vivo* activity[a]
Ser	+
Cys	+
Ala	+
Thr	+/–
Pro	+/–
Asp	–
Asn	–
Glu	–
Gln	–
Gly	–
His	–
Met	–
Val	–
Leu	–

[a] '+' means that UA2F cells expressing Arc with the above position 32 mutants were streptomycin resistant and chloramphenicol sensitive. '–' means that UA2F cells expressing Arc with the above position 32 mutations were chloramphenicol resistant and streptomycin sensitive.

Table 2. *Thermal stability and change in thermal stability of Arc position-32 mutants.*

residue 32	T_m [a]	ΔT_m [a, b]
Asp	71	+7
Ser	64	--
Cys	60	-4
Thr	58	-6
Asn	56	-8
Glu	51	-13
Pro	49	-15
Gly	48	-16
His	46	-18
Ala	45	-19
Gln	43	-21
Met	40	-24
Val	36	-28
Leu	35	-29

[a] T_m and ΔT_m units are °C. [b] ΔTm is the change in thermal stability relative to wild-type Arc.

Table 3. *Protein concentration required for half-maximal binding to the O1 operator.*

residue 32	1/2 maximal concentration
Pro	0.56 nM
Ser (wt)	0.57 nM
Cys	4.5 nM
Asn	5.5 nM
Ala	5.5 nM
Gly	48 nM
Thr	73 nM
Val	79 nM
His	140 nM
Glu	850 nM
Gln	1300 nM
Met	2000 nM
Leu	22,000 nM
Asp	33,000 nM