Relationship between Sequence Determinants of Stability for Two Natural Homologous Proteins with Different Folds†

Laura O. Van Dorn, Tracey Newlove, Saemin Chang, Wendy M. Ingram, and Matthew H. J. Cordes*

Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, Arizona 85721

Received May 1, 2006; Revised Manuscript Received July 1, 2006

ABSTRACT: In the Cro protein family, an evolutionary change in secondary structure has converted an α-helical fold to a mixture of α-helix and β-sheet. P22 Cro and λ Cro represent the ancestral all-α and descendant α+β folds, respectively. The major structural differences between these proteins are at the C-terminal end of the domain (residues 34−56), where two α-helices in P22 Cro align with two β-strands in λ Cro. We sought to assess the possibility that smooth evolutionary transitions could have converted the all-α structure to the α+β structure through sequences that could adopt both folds. First, we used scanning mutagenesis to identify and compare patterns of key stabilizing residues in the C-terminal regions of both P22 Cro and λ Cro. These patterns exhibited little similarity to each other, with structurally important residues in the two proteins most often occurring at different sequence positions. Second, "hybrid scanning" studies, involving replacement of each wild-type residue in P22 Cro with the aligned wild-type residue in λ Cro and vice versa, revealed five or six residues in each protein that strongly destabilized the other. These results suggest that key stability determinants for each Cro fold are quite different and that the P22 Cro sequence strongly favors the all-α structure while the λ Cro sequence strongly favors the α+β structure. Nonetheless, we were able to design a "structurally ambivalent" sequence fragment (SASF1), which corresponded to residues 39−56 and simultaneously incorporated most key stabilizing residues for both P22 Cro and λ Cro. NMR experiments showed SASF1 to stably fold as a β-hairpin when incorporated into the λ Cro sequence but as a pair of α-helices when incorporated into P22 Cro.

Almost 1000 different protein domain folds are known (*1*), but the origin of this structural diversity is, for the most part, a mystery. A protein's fold is encoded in its amino acid sequence and could in principle change to a new fold as this sequence acquires mutations during evolution. Recent comparisons of the structures of homologous proteins have revealed several cases in which new protein folds have derived from evolutionary modifications of preexisting folds (*2*−*4*). We are interested in elucidating the possible mechanisms by which an ancestral protein sequence that specifies an ancestral fold could be mutated to encode a distinct descendant structure.

During evolution, a typical protein undergoes extensive sequence drift with little effect on its native structure and stability. There are two main reasons for this sequence plasticity of proteins. First, a large number of different sequences, many of which share no recognizable similarity, can encode the same fold with high thermodynamic stability. Second, theoretical models of protein evolution suggest that the diverse sequences encoding a particular fold are interconnected as a large "neutral network" in sequence space (*5*−*7*). In such networks, all the constituent sequences are linked by continuous "neutral paths", each of which consists of a series of point mutations or minor insertion−deletion events (indels).[1] The continuity of neutral networks allows for efficient neutral drift in evolution, so a protein may gradually and smoothly accumulate extensive changes to its sequence while retaining the same stable native fold. The orange oval in Figure 1A represents a portion of sequence space corresponding to a hypothetical neutral network for a fold consisting of all α-helical secondary structure (Figure 1C, orange). The mutational trajectory leading from ancestral sequence 1 to descendant sequence 2 is an example of a neutral path within this network.

A fundamental question is whether, during the course of structurally neutral sequence drift, exchange can occur between different neutral networks, leading to evolution of one fold from another. Because different structural topologies impose different requirements on the sequence, the neutral networks for two folds are likely to be quite distinct and may very well have no sequences in common or may exhibit very sparse connections. For example, for the all-α and α+β folds in Figure 1C, the hypothetical networks depicted in Figure 1A (orange and yellow ovals, respectively) do not come close to intersecting in sequence space. In such a case, there is no smooth stability neutral path for evolution of one

† Supported by NIH Grant RO1 GM066806b.

* To whom correspondence should be addressed. Telephone: (520) 626-1175. Fax: (520) 626-9204. E-mail: cordes@email.arizona.edu.

[1] Abbreviations: NMR, nuclear magnetic resonance; CD, circular dichroism; rmsd, root-mean-square deviation; TOCSY, total correlation spectroscopy; NOESY, nuclear Overhauser effect spectroscopy; HSQC, heteronuclear single-quantum coherence; NOESY-HSQC, two-dimensional/three-dimensional heteronuclear NOESY; TOCSY-HSQC, two-dimensional/three-dimensional heteronuclear TOCSY; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; IPTG, isopropyl β-D-thiogalactoside; TMSP, sodium 3-trimethylsilylpropionate; SASF, structurally ambivalent sequence fragment; indel, insertion−deletion mutation.

10.1021/bi060853p CCC: $33.50
Published on Web 08/15/2006

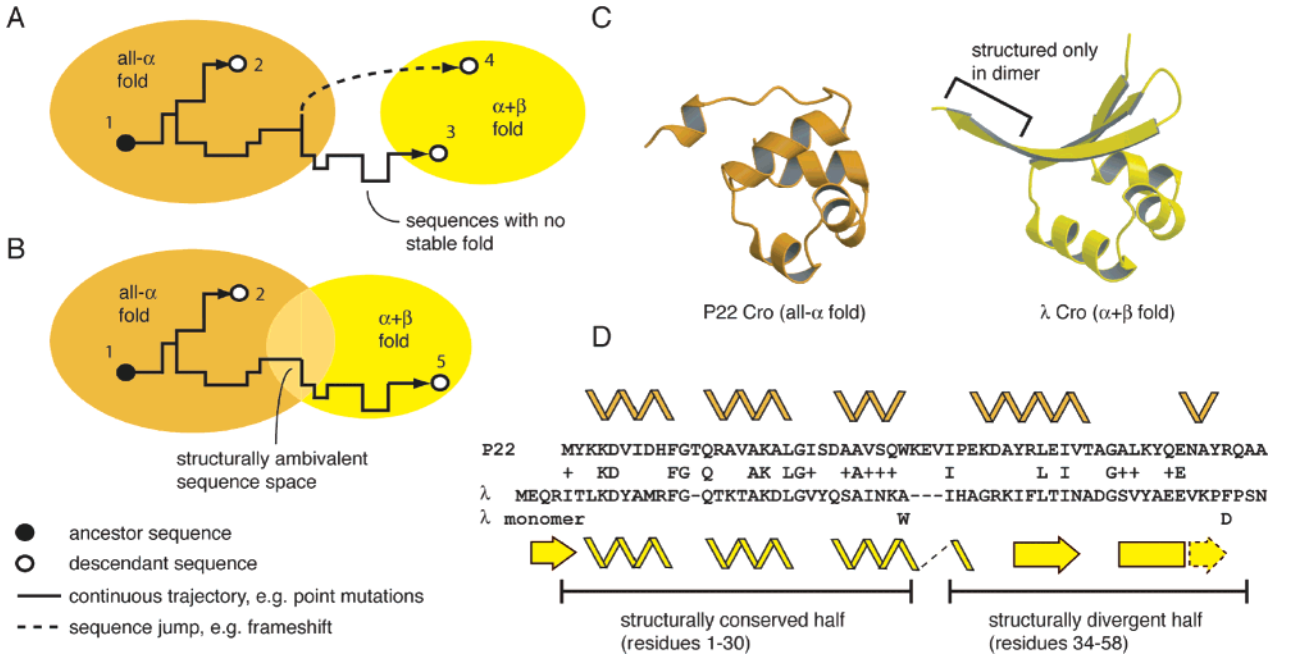

FIGURE 1: Evolution of a new protein fold from a preexisting fold. The orange and yellow ovals in panels A and B crudely represent hypothetical regions of sequence space (neutral networks of sequences) that encode the two folds shown in panel C (see the text), for (A) a case in which no overlap or intersection exists between the two networks and (B) a case in which overlap does exist. The filled and empty circles represent ancestral and descendant sequences, respectively, while solid and dashed arrows represent continuous and discontinuous mutational paths between them, respectively (see the text for an explanation). Panels C and D show a comparison between P22 Cro (orange) and λ Cro (yellow), two related proteins with different folds: (C) ribbon diagrams of each protein, generated using MolScript (*41*) and Raster3d (*42*), with only one subunit of the λ Cro dimer shown, and (D) sequence alignment of P22 Cro and λ Cro, annotated with secondary structures.

fold from the other. Descendants of all-α ancestral sequence 1, for example, cannot acquire the α+β fold by any continuous mutational trajectory that conserves stability. Exchange between networks could still occur, but only if relaxed selection for stability allowed for drift outside the neutral networks (e.g., 1 → 3 trajectory) or if discontinuous mutational events or "sequence jumps" could occur (e.g., 1 → 4 trajectory), for example, by heterologous recombination or frameshifting.

Some theoretical studies have, however, suggested that neutral networks for different folds could closely approach each other or even overlap (*8*, *9*), potentially allowing exchange by continuous mutational paths. In support of this possibility, design and mutagenesis experiments have yielded pairs of similar sequences with different folds (*10*−*13*) and even single sequences that can switch between two folds (*14*−*16*). In the example in Figure 1, if the network for the α+β fold shared sequences in common with that for the all-α fold (Figure 1B), continuous stability neutral paths (e.g., 1 → 5 trajectory) would then exist for interconversion of the two folds, though the number of such paths could be small. The set of sequences in the region of overlap would provide a conduit or channel for fold evolution by smooth mutational mechanisms. These intermediate sequences could be described as structurally ambivalent because they can adopt both folds stably.

Our study focuses on these considerations as they apply to an evolutionary fold change in the Cro family of bacteriophage transcription factors (*4*). The Cro protein from phage P22 (Figure 1C, orange) has a fold consisting of five α-helices, with the second and third forming a helix−turn−helix (HTH) DNA-binding motif. The fold of its ortholog from phage λ shares the first three helices, but the fourth and fifth are replaced by a β-hairpin (Figure 1C, yellow) (*17*). These topological differences map to the C-terminal halves of the sequences, consisting approximately of residues 34−58 (Figure 1D). The N-terminal halves are very similar in structure, except that a short extension at the N-terminus of λ Cro interacts with the C-terminal hairpin to form a three-stranded β-sheet. P22 Cro's all-α fold is shared by distant paralogs of Cro in the CI family, while λ Cro is the only known protein with its particular α+β topology. Hence, the ancestral Cro protein probably had the all-α fold, and this structure was retained in P22 Cro, a conservation which could be represented by the 1 → 2 neutral path (Figure 1A,B). The α+β fold of λ Cro then evolved from it in some subgroup of the family through replacement of α-helical by β-sheet secondary structure. Such a fold change might be represented in panel A or B of Figure 1 by trajectory 1 → 3, 1 → 4, or 1 → 5.

Which of these mutational pathways best describes the evolution of λ Cro's fold is unclear a priori. However, the continuous and discontinuous trajectories can be distinguished by sequence comparisons. The overall sequences of P22 Cro and λ Cro are ~25% identical (Figure 1D). At this high level of divergence, pairwise sequence comparison cannot reveal whether the two Cro proteins and their folds are linked by continuous mutational pathways or by sequence jumps. PSI-BLAST and transitive homology analyses, however, confirmed a distant global sequence homology between P22 Cro and λ Cro that is inconsistent with wholesale nonhomologous replacement of large fragments of the sequence and disfavors the involvement of sequence jumps in Cro evolution (*4*). Instead, the sequences of P22 Cro and λ Cro are probably connected through a long series of substitutions, along with minor indels corresponding to small

gaps in the alignment (Figure 1D). Thus, the pathway from an all-α ancestral Cro to the α+β λ Cro likely proceeded through structurally ambivalent intermediates mediated by neutral network overlap (e.g., 1 → 5 trajectory), or through sequences with no stable fold (e.g., 1 → 3 trajectory), but not through major sequence jumps (e.g., 1 → 4 trajectory).

To begin exploring the plausibility of neutral network overlap and structurally ambivalent sequence intermediates in Cro evolution, we have undertaken a mutagenesis and design study using P22 Cro and λ Cro as respective representatives of the all-α and α+β Cro folds. This work focuses specifically on the structurally diverse C-terminal half of Cro. Our study has three goals: (1) to establish and compare the key C-terminal sequence features critical for the stability of each protein, primarily using alanine scanning mutagenesis, (2) to identify sequence features that confer specificity for one structure over the other, as well as those compatible with both folds, using a technique we call hybrid scanning mutagenesis, and (3) to design structurally ambivalent C-terminal sequence fragments which, when incorporated into Cro sequences, can stably fold into both α-helical and β-sheet conformations.

EXPERIMENTAL PROCEDURES

Cloning, Mutagenesis, and Design Construction. The gene for P22 Cro was tagged with a C-terminal six-histidine sequence and NdeI and BamHI sites using PCR amplification and then ligated into the pET21b vector to give the pTN200 expression plasmid. In a similar manner, the λ Cro gene was cloned previously into pET21b using NdeI and XhoI sites, leading to the incorporation of a C-terminal six-histidine tag from the vector, connected to the Cro sequence by a Leu-Glu dipeptide corresponding to the XhoI site (*18*). Plasmids encoding monomeric variants of λ Cro used in this study (pKL105 for A33W/F58D and pTN620 for A33W/F58D/Y26Q) were generated from this parent λ Cro construct (pMC140) using the QuikChange mutagenesis protocol (Stratagene). Alanine and hybrid variants were then constructed from pTN200 or pKL105 using the QuikChange protocol. Structurally ambivalent sequence fragment (SASF) designs were introduced into λ Cro and P22 Cro constructs in parallel using a combination of cassette mutagenesis and QuikChange site-directed mutagenesis using the QuikChange protocol. During introduction of SASF-encoding sequences into the P22 Cro gene, we switched from our original expression construct to one comparable to that used for λ Cro, with the coding sequence flanked by NdeI and XhoI sites, and the histidine tag contributed by the pET21b vector.

Expression and Purification of Proteins. Cro variants were expressed in *Escherichia coli* BL21(λDE3) cells transformed with the appropriate expression plasmid and purified by Ni^{2+}−NTA chromatography in 6 M guanidine essentially as described previously, using 10 mM imidazole in all load and wash steps to reduce nonspecific binding (*18, 19*). Purified proteins were refolded by dialysis into SB250 buffer [50 mM Tris (pH 7.5), 250 mM KCl, and 0.2 mM EDTA] and centrifuged at 12000*g* for 20 min to remove precipitates. In the case of uniform ^{15}N-labeled NMR samples of SASF-containing proteins, expression was performed in M9T minimal medium using 0.8 g/L $^{15}NH_4Cl$ as the sole nitrogen source, and an additional size exclusion purification step was employed. Nickel affinity purified proteins dialyzed into SB250 buffer were loaded onto a HiPrep 26/60 Sephacryl S-100 column (Pharmacia Biotech) equilibrated in the same buffer and eluted isocratically at a flow rate of 1.3 mL/min. Protein-containing fractions, as identified by SDS−PAGE, were combined, dialyzed into NMR buffer [50 mM sodium phosphate (pH 6.3) and 150 mM NaCl], and concentrated to approximately 0.75 mM using Centricon YM-3 or Centriprep YM-3 (Millipore) centrifugal filter devices.

Circular Dichroism Spectroscopy. Far-ultraviolet CD spectra and melts of Cro variants were obtained on Aviv 62A DS and Jasco J-810 CD spectrometers. The two spectrometers yielded essentially identical thermal denaturation midpoints for the parent P22 Cro and λ Cro proteins, verifying that no significant systematic errors were introduced by characterizing proteins on one or the other. Thermal denaturation curves (5 μM protein in SB250) were typically collected from 15 to 80 °C in 1 °C intervals with 1 min equilibration times at each temperature. Reversed melts showed no apparent hysteresis and yielded >90% recovery of the folded baseline signal, verifying that the Cro variants folded and unfolded reversibly. T_m values were obtained by fitting with the Gibbs−Helmholtz equation essentially as described previously (*19*), using Kaleidagraph (Synergy). In these fits, the value of the heat capacity of unfolding (ΔC_p) was fixed at an estimated value of 840 cal mol^{-1} K^{-1} for both P22 Cro and λ Cro variants. This value was computed for a typical 60-residue protein based on an average per-residue ΔC_p of 14 cal mol^{-1} K^{-1} $residue^{-1}$ (*20*). All other parameters, including unfolded and folded baseline slopes and intercepts, ΔH_u, and T_m, were allowed to vary. Variants destabilized by more than ∼15 °C often lacked clear folded baselines, particularly in the case of λ Cro. In such cases, the output T_m was sensitive to the initial guesses of parameter values. Initial guesses of T_m for these variants were generated from overlays of parent protein and variant melts, assuming that the true baselines of the variants resembled those of the parent protein. Initial guesses for baseline slopes and intercepts were based on those of the parent protein. Initial guesses for ΔH_u were related to the initial T_m guess using a plot of ΔH_u versus T_m for a series of variants. These parameters were then allowed to vary during fitting, but fits in which ΔH_u and baseline parameters deviated strongly from the initial guesses were rejected provided that equally good or better fits with smaller deviations could be generated by slightly varying the initial guesses. Highly unstable, mostly unfolded variants for which the initial T_m guess was well below 290 K were not fit at all, and ΔT_m values were conservatively estimated to be no smaller than −35 °C. Finally, the G48A variant of λ Cro exhibited an unusually broad melt that was resistant to convincing fitting. The ΔT_m value reported for G48A in Figure 2 was estimated as an average of the best fit allowing all parameters to vary and a fit in which baseline parameters were restricted to those of the G37A variant. Typical uncertainties in reported ΔT_m values are ±2 °C, increasing to ±5 °C for variants destabilized by more than ∼15 °C or for the G48A variant of λ Cro.

NMR Spectroscopy. NMR spectra were recorded on a Varian Inova 600 MHz spectrometer at 293 K. Samples of P22-SASF1 and λ_{WDQ}-SASF1 contained 0.75 mM uniformly ^{15}N-labeled protein in 50 mM sodium phosphate (pH 6.3),

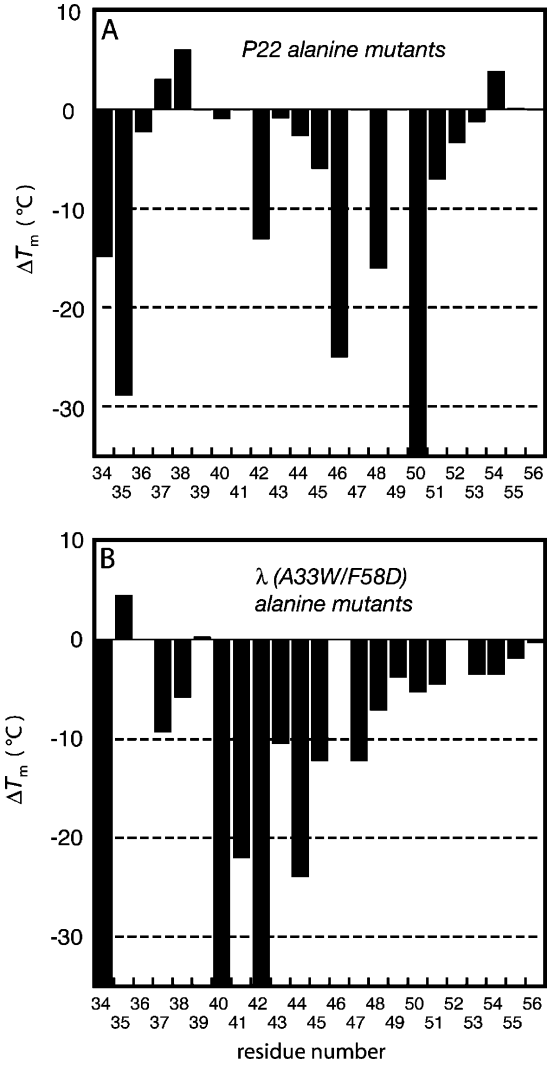

FIGURE 2: Thermal stabilities of alanine point mutants for residues 34−56 of (A) P22 Cro and (B) λ Cro-A33W/F58D, relative to stabilities of the parent proteins. ΔT_m values reported here as −35 °C represent conservative, minimal estimates of destabilization for very unstable variants which were mostly unfolded at ambient temperature (see Experimental Procedures).

150 mM sodium chloride, 10% ^2H_2O, 0.01% sodium azide, and 1 mM TMSP. Spectra were processed using NMRPipe/NMRDraw (*21*) and analyzed using NMRView (*22*). $^{15}N-^1H$ HSQC spectra were obtained for P22-SASF1 and λ_{WDQ}-SASF1 (Figure 6) and were assigned using strip plot analysis of three-dimensional (3D) $^{15}N-^1H$ TOCSY-HSQC (50 ms mixing time) and 3D $^{15}N-^1H$ NOESY-HSQC (50 ms mixing time) spectra. Both spectra were assigned for essentially all residues, though Figure 6 shows resonance assignments for only the SASF region. The known chemical shifts for the parent proteins aided this process considerably (*4*, *23*). H_α chemical shifts (Figure 7) were obtained from the same analysis. Spectra and chemical shifts in Figures 6 and 7 were referenced using the H_2O signal as a standard. TMSP-referenced shifts were considered untrustworthy due to possible transient interactions between the standard and the protein reported previously for λ_{WDQ} (*23*). For P22-SASF1 and λ_{WDQ}-SASF1, a value of 4.811 was used for the H_2O resonance, based on salt and temperature dependencies listed in Wishart et al. (*24*). For the parent proteins P22 Cro and λ_{WDQ} Cro, H_α assignments were obtained in previous work on samples at 293 K, without any salt, and at pH 6.4 and 5.3, respectively. For the comparison in Figure 6, both sets of shifts were referenced to H_2O at 4.826, again based on ref *24*. H_α chemical shifts were adjusted for random-coil values as listed in Wüthrich (*25*).

RESULTS

C-Terminal Sequence Determinants of Stability for P22 Cro and λ Cro. We performed two types of scanning mutagenesis on the C-terminal halves of both P22 Cro and λ Cro. First, we generated a complete set of alanine point mutations and used the effect of each substitution on the folding stabilities of P22 Cro and λ Cro to identify residues important for stably encoding their respective structures (*26*, *27*). Second, we conducted "hybrid scans" in which each position in P22 Cro was substituted with the aligned residue in λ Cro and vice versa, according to the PSI-BLAST alignment in Figure 1D. The hybrid scans supplement the alanine scans as indicators of the mutational sensitivity of particular residues and, thus, may suggest additional stability determinants. The primary purpose of the hybrid scans, covered in a later section, is to evaluate the compatibility of the P22 Cro and λ Cro sequences with the fold adopted by the other protein.

A potential complicating factor in these studies is the fact that P22 Cro and λ Cro differ in oligomeric state as well as fold. λ Cro does not fold stably as a monomer ($T_m \leq 40$ °C) (*28*) and dimerizes in solution at low micromolar concentrations (*18*, *29*, *30*), while P22 Cro is stable as a monomer ($T_m = 55$ °C) and does not dimerize measurably in solution (*4*). The strong dimerization and weak monomer stability in λ Cro are related to a "ball-and-socket" motif (*31*) that unites part of the subunit core with the dimer interface and partly couples folding stability to self-association. Mutational effects on λ Cro stability could therefore include components both from destabilization of the core α+β domain fold and from disruption of the dimer interface.

The folding stability of λ Cro can, however, be uncoupled from dimerization by mutations in the ball and socket (*18*, *23*). The λ Cro double mutant A33W/F58D (Figure 1D) is purely monomeric, with a thermal stability ($T_m \sim 52$ °C) similar to that of the P22 Cro monomer (*18*). The A33W/F58D monomer retains essentially the same structure as individual subunits of the wild-type λ Cro dimer, except that the last two or three residues of strand 3 (approximately residues 56−58), which participate in the dimer interface but make no long-range contacts in the monomer, are disordered (see panels C and D of Figure 1) (*23*). To focus our mutagenesis study on a comparison between determinants of the monomeric core α+β and all-α domain structures, rather than including the complicating issue of dimerization, we used λ Cro-A33W/F58D as a background sequence rather than wild-type λ Cro.

Folding stabilities of alanine and hybrid mutants for residues 34−56 of P22 Cro and λ Cro-A33W/F58D were characterized using reversible thermal denaturation monitored by circular dichroism (CD). The thermal denaturation midpoints of the mutants (T_m) were compared to those of the parent sequences, and the differences (ΔT_m) were plotted in Figures 2 and 3. In interpreting these data below, we

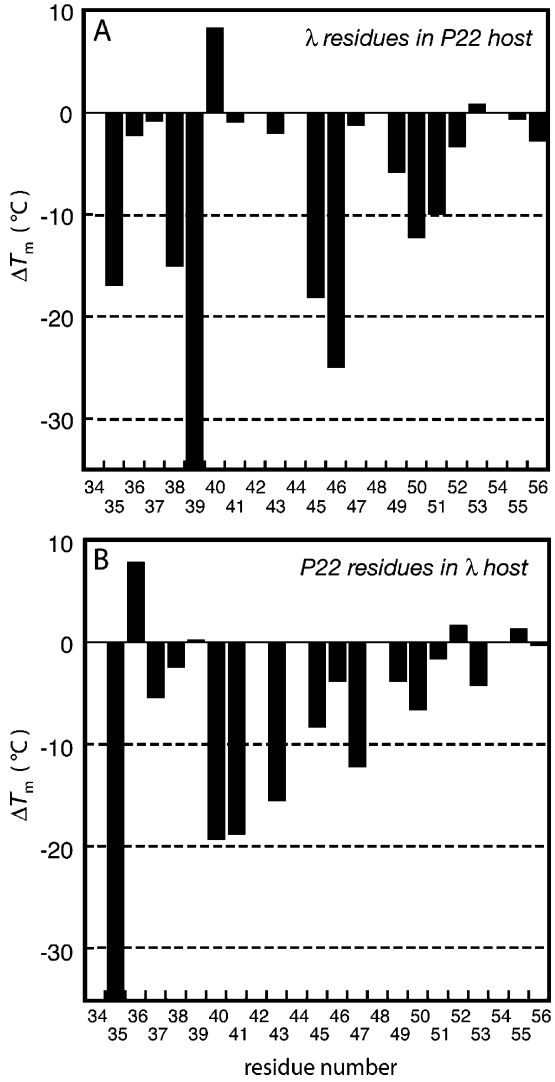

FIGURE 3: Thermal stabilities of hybrid point mutants for residues 34−56 of (A) P22 Cro and (B) λ Cro-A33W/F58D, relative to stabilities of the parent proteins. Hybrid point mutants were generated by introducing a wild-type residue of P22 Cro at the aligned position of a λ Cro-A33W/F58D host sequence or by introducing a wild-type residue of λ Cro into a P22 Cro host sequence. ΔT_m values reported here as −35 °C represent conservative, minimal estimates of destabilization for very unstable variants which were mostly unfolded at ambient temperature (see Experimental Procedures).

assumed that no qualitative native state structural changes were induced by the point mutations and that the changes in T_m directly reflected effects on the equilibrium between a wild-type-like native folded state and an unfolded ensemble. Later in this study, we will present direct evidence that many residues from P22 Cro can be introduced simultaneously into λ Cro and vice versa without qualitatively modifying the structure of either protein, even though many of these residues strongly stabilize alternate native structures. Thus, the structural specificity of these proteins is probably not compromised by simple point mutations.

A high negative ΔT_m value (arbitrarily chosen as greater than −10 °C) in the alanine scans (Figure 2) was thus interpreted as an approximate indication that the deleted side chain was important for the stability of the mutated protein in its native fold. Six to eight residues from each protein exhibited marked destabilization when substituted with alanine. Reductions in stability in the hybrid mutations (Figure 3) were considered supplemental indicators of a residue's importance, accounting for the fact that alanine scans cannot identify all key stability determinants (for example, in situations where alanine itself plays an important structural role). Five to six residues in each protein exhibited marked destabilization when replaced in the hybrid scans, many of which coincided with stability determinants identified in the alanine scans. For each protein, a total of nine residues exhibited ΔT_m values of greater than −10 °C in alanine and/or hybrid scans. These positions were mapped onto the structures and sequences of P22 Cro and λ Cro to illustrate differences and similarities (Figure 4). These sets of mutation-sensitive positions are clearly quite different (Figure 4E) and correlate with distinct interactions present in the known structures of the two proteins (Figure 4A−D).

Structural Roles of Key Stability Determinants. In P22 Cro, the mutation-sensitive residues are scattered across the sequence (Figure 4E) but are clustered in the structure (Figure 4A), with most facing the interior of the all-α fold. Ile 34, Asp 38, Ala 39, Leu 42, Thr 46, and Leu 50 all have less than 10% solvent accessible surface area and might be considered core residues. Thr 46 and Leu 50 exhibit severe intolerance to mutation, with destabilization observed in both alanine and hybrid scans. Thr 46 not only is buried but also makes a side chain−main chain hydrogen bond to residue 49 or 50. Ile 34, Ala 39, and Leu 42 cannot be studied by both techniques since they either are Ala in the wild-type sequence (Ala 39) or are identical in λ and P22 Cro (Ile 34 and Leu 42). Asp 38, interestingly, tolerates substitution to Ala quite well but not substitution with the wild-type λ Cro residue Arg. Asp 38 probably accepts a hydrogen bond from another largely buried side chain, Trp 33 (Figure 4B). Mutation of Asp 38 to Ala removes the acceptor carboxylate group but may permit entry of solvent to hydrogen bond to Trp 33 in its place. Mutation to Arg also removes the carboxylate, but the presence of a γ-methylene group may block solvent access to Trp 33.

The remaining mutation-sensitive residues in P22 Cro either are on the periphery of the core (Pro 35 and Val 45) or participate in turns and helix termination (Pro 35 and Gly 48). Val 45 is weakly sensitive to Ala substitution but does not tolerate polar residues well, as judged by the destabilization observed in the Val45Asn hybrid variant. The N-terminus of helix 4 has some of the sequence characteristics of a Pro-box capping motif as described by Viguera and Serrano (*32*), with Pro 35 occupying the N-cap position. Helix N-capping may therefore explain the sensitivity of Pro 35 to alanine and hybrid substitutions. Gly 48, though not part of any canonical helix C-capping motif (*33*, *34*), occupies position $i + 2$ of a type I′ β-turn immediately following helix 4. Glycine is strongly preferred at such positions in natural proteins (*35*).

In λ Cro, mutation-sensitive residues are concentrated in a contiguous stretch of sequence (Figure 4E) corresponding to the central strand of the three-stranded β-sheet (Figure 4C). As in P22 Cro, the mutation-sensitive residues in λ Cro are still mostly shielded from solvent, but they are not confined to the interior face of the C-terminal secondary structural elements. Three of the nine sensitive residues have less than 10% accessible surface area (Ile 40, Leu 42, and Thr 43), while three more (Ile 34, Phe 41, and Ile 44) have

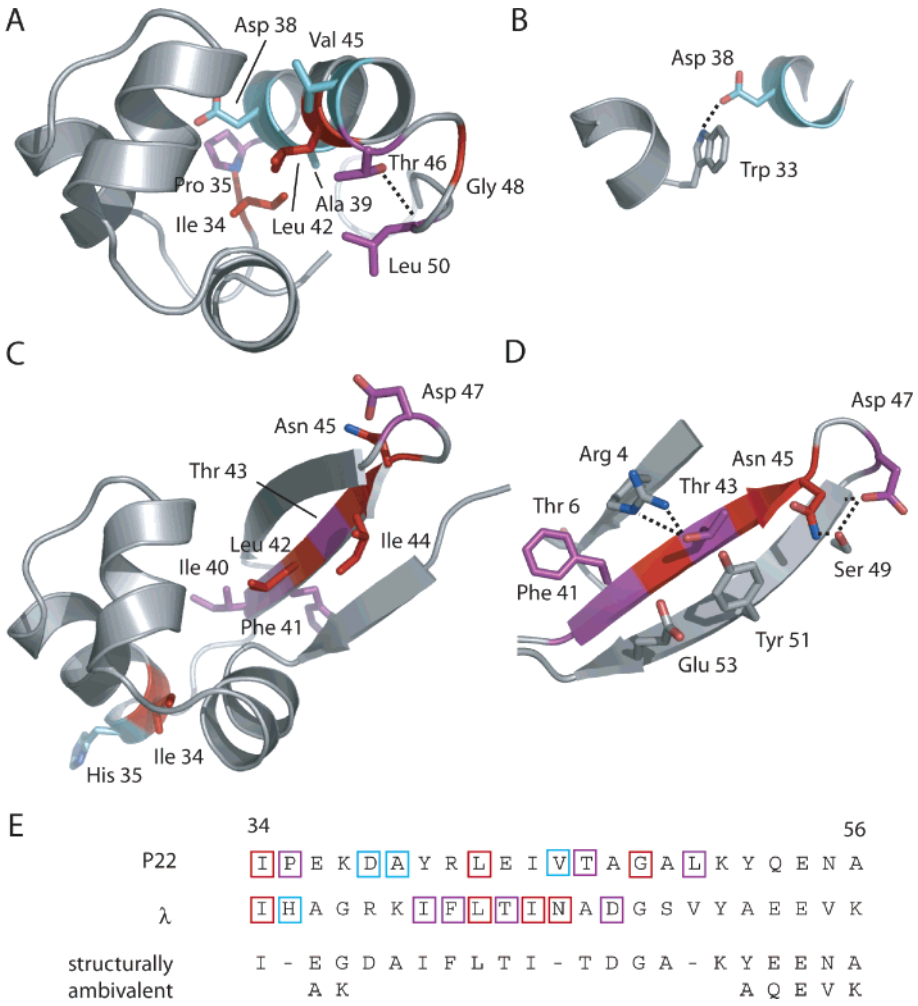

FIGURE 4: Mutation-sensitive positions in P22 Cro and λ Cro-A33W/F58D, showing large decreases in stability upon substitution with alanine (red), hybrid mutation (blue), or both (purple): (A) mutation-sensitive residues in P22 Cro mapped onto the structure (PDB entry 1RZS) (*4*), (B) picture of Asp 38−Trp 33 interaction in P22 Cro, (C) mutation-sensitive residues in λ Cro mapped onto one subunit (PDB entry 5CRO) (*17*), (D) view of the solvent-facing side of λ Cro's β-sheet, with dashed lines indicating hydrogen bonds, and (E) mutation-sensitive residues mapped onto a sequence alignment of P22 Cro and λ Cro for the region of residues 34−56. Also shown are "structurally ambivalent" residues at each position identified through hybrid scanning mutagenesis (see the text).

10−15%. Of this group of six, four clearly face the protein interior (Ile 34, Ile 40, Leu 42, and Ile 44). Phe 41 and Thr 43, on the other hand, are on the solvent-facing side of the β-sheet but pack against other side chains on the same strand and on flanking strands (Figure 4D). Thr 43 also makes a cross-strand hydrogen bond to Arg 4. All six of these buried or partly buried residues either are sensitive in both alanine and hybrid scans or cannot be tested in the hybrid scans because they are identical in λ Cro and P22 Cro.

The three remaining mutation-sensitive residues in λ Cro are solvent-exposed. Asn 45 and Asp 47 form part of a network of hydrogen bonds in the turn between β-strands 2 and 3 (Figure 4D). His 35 is a special case, because it probably does not contribute much to λ Cro stability per se, being on the surface of the end of helix 3 and making no apparent interactions with other side chains. This residue is quite tolerant to alanine substitution. However, P22 Cro has a proline at this position, which would be expected to disrupt helix 3 in λ Cro.

Comparison of Stability Determinant Patterns in P22 Cro and λ Cro. A comparison of the key stabilizing residues in P22 Cro and λ Cro can, to a first approximation, identify similarities and differences in the way the all-α and α+β Cro folds are encoded in the sequence and permit a rough assessment of the potential for the neutral networks of the two folds to intersect. First, we examined four C-terminal positions (34, 35, 42, and 45; Figure 4E) which exhibited sensitivity to substitution in both P22 Cro and λ Cro. Two of these positions (34 and 42) involved the same residue type playing essentially the same structural role in both proteins. Positions 34 and 42 are occupied by Ile and Leu, respectively, in both P22 Cro and λ Cro and form part of each protein's hydrophobic core. Such positive overlap between key stability determinants for two folds implies similarities in the way they are encoded and may increase the likelihood that their neutral networks have sequences in common. Positions 35 and 45, however, exhibited specific sequence requirements that directly conflicted in P22 Cro and λ Cro. Position 35 involved an important Pro residue in P22 Cro which was quite sensitive to substitution. His 35 of λ Cro, while not important in and of itself, also did not tolerate substitution with Pro due to its location in the middle of a helix. Position 45 is a partly buried Val in P22 Cro but in λ Cro is an Asn that makes solvent-exposed hydrogen bonding interactions. Such conflict or negative overlap between key determinants for two folds decreases the

Table 1: Host Sequences, SASF Designs, and SASF Constructs Used in This Study[a]

```
residue #                                          .35...40...45...50...55...60
λ ss          eee hhhhhhhh   hhhhhhh   hhhhhhh---hhh   eeeeee    eeeeee
p22 ss            hhhhhhhh   hhhhhhh   hhhhhh      hhhhhhhhhh       hhhh

λWDQ          MEQRITLKDYAMRFG-QTKTAKDLGVQQSAINKW---IHAGRKIFLTINADGSVYAEEVKPDPSN
p22               MYKKDVIDHFGTQRAVAKALGISDAAVSQWKEVIPEKDAYRLEIVTAGALKYQENAYRQAA

λWDQ-SASF1    MEQRITLKDYAMRFG-QTKTAKDLGVQQSAINKW---IHAGRAIFLTIYTDGRLKMEENAYRQAA
p22-SASF1         MYKKDVIDHFGTQRAVAKALGISDAAVSQWKEVIPEKDAIFLTIYTDGRLKMEENAYRQAA
                     **    ** *   **  *    *   *   *    **********************

λWDQ-SASF2    MEQRITLKDYAMRFG-QTKTAKDLGVQQSAINKW---IHEKDAIFLTIYTDGRLKMEENAYRQAA
p22-SASF2         MYKKDVIDHFGTQRAVAKALGISDAAVSQWKEVIPEKDAIFLTIYTDGRLKMEENAYRQAA
                     **    ** *   **  *    *   *   * **************************
```

[a] All SASF constructs contained a C-terminal LEHHHHHH tag. Residues 62−66 of λ_{WDQ} and λ_{WDQ}-SASF1 (KKTTA) are not shown. Italicized residues in SASF1 and SASF2 are in a disordered region of either P22 Cro, λ Cro, or both. Asterisks denote identical residues. SASF regions are underlined.

likelihood of overlap between their neutral networks and, if severe enough, could completely prevent it.

Most mutation-sensitive positions identified in our study, however, involved neither positive nor negative overlap between stability determinants of the two structures. Specifically, 10 C-terminal positions exhibited structural importance in only one protein or the other (Figure 4E). The prevalence of such positions suggests a certain orthogonality or independence between key C-terminal stability determinants of each of the two Cro folds. From the perspective of overlap between their neutral networks, this orthogonality poses an interesting dilemma. Completely orthogonal or nonoverlapping sets of determinants would allow both folds to be encoded in a single sequence but would also restrict such structurally ambivalent sequences (Figure 1B) to the small number that contain both sets of determinants simultaneously, creating an evolutionary "bottleneck". A more detailed treatment of this idea is left to the Discussion.

Specificity Determinants and Structurally Ambivalent Residues in P22 Cro and λ Cro. The apparent differences in stability determinants for the two Cro folds imply that some sequence features might strongly favor one fold over the other. The hybrid scanning experiments reveal residues in the sequences of P22 Cro and λ Cro which are compatible or incompatible with the sequence or structure context of the other protein, and thus to a certain extent with an alternate native Cro fold. For example, a residue from α+β λ Cro which causes minimal destabilization when introduced into the all-α P22 Cro sequence can be regarded as compatible with both folds and, thus, as structurally ambivalent with respect to Cro secondary structure switching. On the other hand, one which causes destabilization of P22 Cro might be a determinant of structural specificity (*36*), favoring the α+β fold over the all-α fold.

The C-terminal regions of P22 Cro and λ Cro both contain five or six residues which strongly destabilized the structure of the other protein ($\Delta T_m > -10$ °C) (Figure 3) and might therefore be features supporting structural specificity. Here we discuss a few illustrative examples. In λ Cro, the large polar charged residue Lys 39 corresponds to a position occupied by a small nonpolar buried residue (Ala 39) in P22 Cro. Ala 46 in λ Cro cannot substitute for the role of Thr 46 in P22 Cro,which makes a key buried side chain−main chain hydrogen bond. Pro 35 of P22 Cro aligns to a position in the middle of an α-helix in λ Cro and thus disrupts its structure. Tyr 40 of P22 Cro changes the shape and volume of a critical hydrophobic core residue in λ Cro. Ala 47 in P22 Cro cannot substitute for a key Asp residue in λ Cro that makes hydrogen bonds in a hairpin turn.

These findings suggest that at least a few substitutions would be necessary to convert the C-terminal region of either protein into a sequence compatible with the alternate Cro fold. In other words, the sequence of λ Cro appears to have drifted quite significantly away from its all-α ancestry, while that of P22 Cro does not appear poised to switch from the ancestral all-α fold to the α+β fold. Viewed in terms of the sequence space diagrams of Figure 1, both P22 Cro and λ Cro are probably located some distance from any structurally ambivalent region of sequence space.

At the same time, most C-terminal residues in each protein do appear structurally ambivalent in the sense of basic compatibility with both Cro folds. Sixteen of 23 P22 Cro residues in the region of residues 34−56 reduced the T_m of λ Cro by less than 5 °C, and 15 of 23 λ Cro residues reduced the T_m of P22 Cro by less than 5 °C (Figure 3). These structurally ambivalent residues are listed in Figure 4E. At 20 of 23 positions, one or both of the P22 Cro and λ Cro wild-type residues were structurally ambivalent by this measure. At only one position, 35, were both hybrid mutants severely destabilizing ($\Delta T_m > -10$ °C) such that neither wild-type residue showed any significant compatibility with both structures. These findings raise the possibility that, even though the C-terminal sequences of P22 Cro and λ Cro as a whole appear to be quite structurally specific, some simple hybrid of these sequences might be close to being structurally ambivalent.

Design of Structurally Ambivalent Cro C-Terminal Sequence Fragments. To test this idea, we used the ambivalent residues shown in Figure 4E as the primary basis for the design of structurally ambivalent sequence fragments (SASFs) or "chameleon sequences" corresponding to most of the C-terminal half of Cro (Table 1). In these designs, we generally incorporated the most ambivalent of the P22 Cro and λ Cro wild-type residues at each position as identified through hybrid scans (see Figure 4E). To a great extent, this approach also has the effect of incorporating the key

stabilizing residues from each protein. In the course of preliminary testing of designs, several adjustments were made, including incorporation of residue types not present in either P22 Cro or λ Cro at three positions (Tyr 45, Arg 49, and Met 52). The incorporation of Tyr at position 45 and Arg at position 49 was based on a heuristic analysis of Cro sequence alignments (not shown). The use of Met 52 was based on previous studies showing that this residue stabilized monomeric versions of λ Cro by 5−7 °C relative to Ala 52 of λ Cro and Tyr 52 of P22 Cro (*23*). We also tried to find residues at position 35 which were more structurally ambivalent than either His 35 of λ Cro or Pro 35 of P22 Cro (see Figure 3). These efforts were unsuccessful, and since both of these residues severely destabilized one Cro structure or the other, we abandoned attempts to incorporate residue 35 into our chameleons and concentrated on the remainder of the C-terminal region.

The properties of the designs were tested by introducing them into both P22 and λ host sequences (see Table 1) using a combination of cassette and site-directed mutagenesis. Here we report results for two designs of slightly different length: SASF1, which spans residues 39−56, and SASF2, which spans residues 36−56 (see Table 1). In constructing these designs, we also replaced residues 57−66 of λ Cro, which are unstructured in the λ Cro monomer, with residues 57−61 of P22 Cro (see Table 1). For the P22 host background, we used the wild-type sequence, while for the λ host sequence, we utilized a Y26Q variant of λ Cro-A33W/F58D (λ_{WDQ} for short), which increases its thermal stability by 9 °C but does not perturb its structure significantly (*23*).

Thermal denaturation curves showed that both SASF1-containing proteins were slightly less stable than their parent sequences (Figure 5). λ_{WDQ}-SASF1 is 5 °C less stable than λ_{WDQ}, with a T_m of 56 °C (Figure 5B), while P22-SASF1 was destabilized by 8 °C relative to P22 Cro and had a T_m of 46 °C (Figure 5A). Both proteins, however, exhibited clear folded baselines and exhibited fully reversible unfolding. The SASF2-containing proteins also appeared to be folded (Figure 5A,B) and gave reversible melts. P22-SASF2 has the same primary sequence as P22-SASF1, because the extension of the chameleon to include residues 36−38 incorporated the three wild-type P22 Cro residues for this tripeptide (EKD; see Table 1). λ_{WDQ}-SASF2 is destabilized by 8 °C relative to λ_{WDQ}-SASF1 due to incorporation of this EKD sequence in place of the wild-type AGR tripeptide. These results demonstrate the replacement of almost the entire C-terminal half of P22 Cro and λ Cro in parallel with the same designed sequence fragment, without catastrophic destabilization of either protein.

Both λ_{WDQ}-SASF1 and P22-SASF1 exhibited sufficient solubility for extensive NMR characterization. ^{15}N-labeled samples of each protein at 0.75 mM showed well-dispersed ^{15}N−^{1}H correlation spectra, consistent with specific folded structures (Figure 6). Sequence-specific resonance assignment of these spectra using 3D isotope-edited NOESY and TOCSY experiments showed that the ^{15}N and ^{1}HN resonance positions of the SASF1 region (residues 39−56) differed strongly in P22-SASF1 and λ_{WDQ}-SASF1 (Figure 6), though the primary sequence of this region is exactly the same in both cases. A number of residues have ^{1}HN chemical shifts which differ by at least ∼1 ppm between the two host contexts. These extreme context-dependent changes suggest

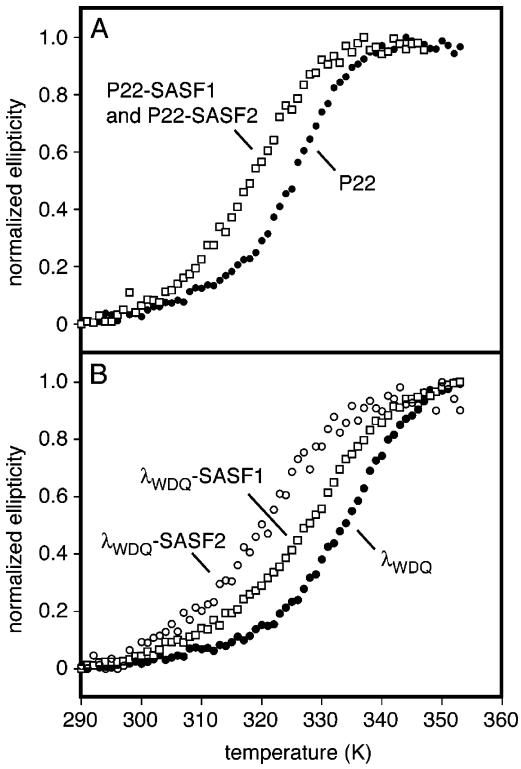

FIGURE 5: Thermal denaturation curves, monitored by circular dichroism at 222 nm, for (A) P22 and (B) λ_{WDQ} with and without designed structurally ambivalent sequence fragments (SASF1 or SASF2) inserted in the C-terminal region. This figure shows that the SASFs destabilize both proteins, but not severely.

that the SASF1 chameleon is adopting very distinct conformations depending on whether it is introduced into the P22 or λ_{WDQ} host sequence.

Analysis of H_α chemical shifts showed that the different structures adopted by SASF1 in P22 and λ_{WDQ} corresponded to those adopted by the respective parent sequences (Figure 7). Random-coil adjusted H_α shifts for residues 39−56 in λ_{WDQ} and λ_{WDQ}-SASF1 are very similar to each other, as are those for P22 and P22-SASF1. Both λ_{WDQ} and λ_{WDQ}-SASF1 have two stretches (residues 40−45 and 50−54) with high H_α chemical shifts, corresponding to β-strands 2 and 3 in λ_{WDQ}, respectively. These stretches are separated by the hairpin turn (residues 46−49), which again shows similar shifts in both the λ_{WDQ} parent and λ_{WDQ}-SASF1. P22 and P22-SASF1 are also similar, with low H_α chemical shifts essentially across the board, consistent with α-helical structure. By contrast, the patterns for the two P22-based sequences differ strongly from those for the λ-based sequences. In particular, P22-SASF1 and λ_{WDQ}-SASF1 have very different H_α shifts despite having exactly the same primary sequence in this region. Thus, SASF1 is a true structurally ambivalent sequence fragment or chameleon: it conforms to a predominantly α-helical structure when introduced into P22 Cro but to a β-sheet structure when introduced into λ_{WDQ}.

We were not able to complete a similar NMR analysis of SASF2. As noted above, the sequence of P22-SASF2 is identical to that of P22-SASF1, so the analyses just presented already verify a helical structure for SASF2 in the P22 sequence context. However, insufficient long-term solubility of λ_{WDQ}-SASF2 at high concentrations in NMR samples

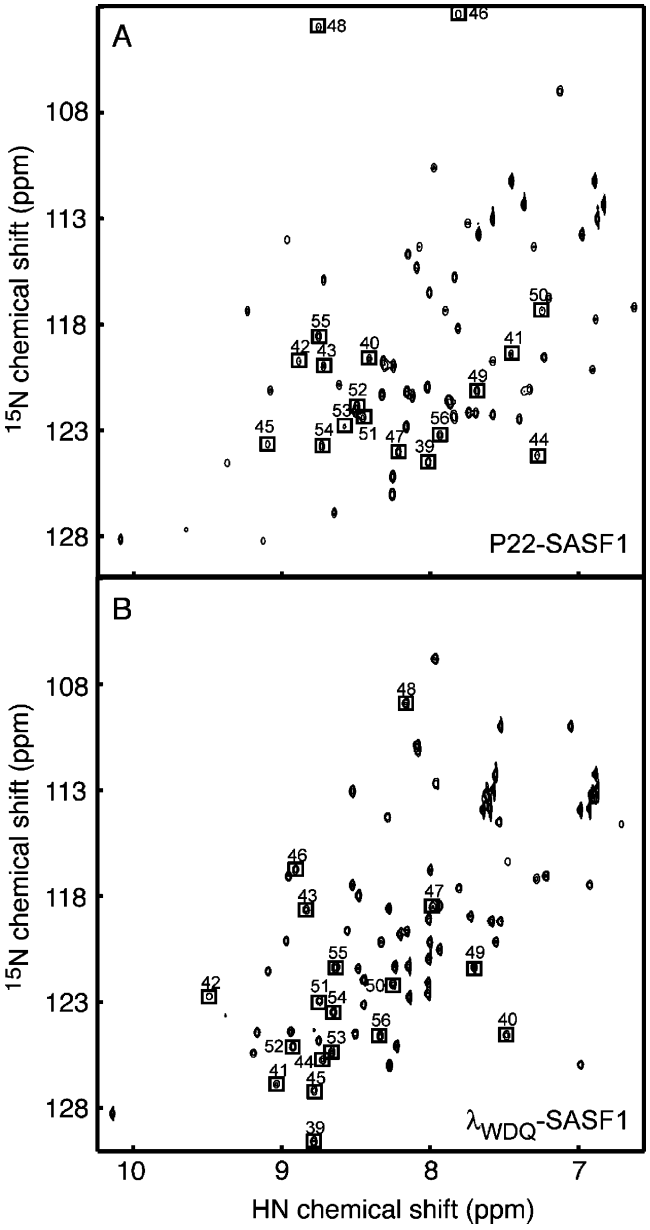

FIGURE 6: ^{15}N−^{1}H HSQC spectra of (A) P22-SASF1 and (B) λ_{WDQ}-SASF1, showing assigned peaks for the SASF region. This figure shows that the SASF has a different structure in the two host sequences, despite having the same primary sequence.

prevented confident resonance assignments for this protein. λ_{WDQ}-SASF2 did remain in solution for sufficient time to yield a well-dispersed HSQC spectrum qualitatively similar to that of λ_{WDQ}-SASF1 (Figure S-1), suggesting a folded, nativelike protein with some β-sheet structure in the SASF region. In particular, the HN and ^{15}N resonance positions of several fairly well-resolved residues within the SASF region, including Ile 40, Phe 41, Thr 43, and Gly 48, were tentatively identified and were not strongly different from the corresponding resonance positions in λ_{WDQ}-SASF1. At the same time, the overall spectral overlay of λ_{WDQ}-SASF2 and λ_{WDQ}-SASF1 does exhibit some significant changes in resonance positions. Though SASF2 probably acts as a 21-residue chameleon corresponding to the entirety of helices 4 and 5 of P22 Cro as well as the β-hairpin of λ Cro (residues 36−56), its secondary structure in the λ_{WDQ} context cannot be assigned with confidence at present.

Although a full C-terminal SASF design (residues 34−56) has not been completed, our results demonstrate that chameleon sequences corresponding to the great majority of the structurally divergent region of Cro are not particularly difficult to construct. The SASF design was, for the most part, a simple hybrid of the P22 and λ C-terminal sequences with a small number of modifications. Important C-terminal determinants for each Cro structure, with the possible exception of those at position 35, can be incorporated simultaneously in a single primary sequence.

DISCUSSION

In this study, we used mutagenesis and design to conduct a limited exploration of the sequence space encoding two different folds that are related by an evolutionary secondary structure switching event. We used P22 Cro and λ Cro as respective representatives of ancestral all-α and descendant $\alpha+\beta$ folds in the Cro family of transcription factors. Our study focused on the C-terminal region of each sequence, which contains the major structural differences between the two folds. Our major findings are that (1) the key C-terminal sequence features responsible for encoding the structure of each protein neither reinforce each other nor conflict strongly but are instead mostly orthogonal, occurring at different positions in the sequence, and (2) single structurally ambivalent or chameleon sequences representing most of the C-terminal region can be designed that stably adopt a helical or sheet conformation depending upon the global sequence of the Cro protein into which they are introduced. From these results, we draw the general conclusion that overlap or near overlap between the neutral sequence networks encoding different Cro structures is probable (Figure 1B), but the number of structurally ambivalent sequences (those compatible with both folds) will be limited by the necessity to simultaneously satisfy two mostly independent sets of sequence requirements. In other words, continuous stability-conserving mutational pathways (e.g., 1 → 5 trajectory in Figure 1B) between ancestral and descendant Cro folds could exist but might traverse a narrow region of overlap, or a bottleneck, in sequence space.

At this point, we have little basis for estimating the intrinsic likelihood or probability of smooth evolutionary exchange between the two Cro folds relative to any other given pair of folds. It seems inevitable that any pair of structures with major topological differences, especially those involving changes in secondary structure type, will have very different patterns of stability determinants in the sequence. The exchangeability of two folds by continuous pathways in evolution will be a function not only of positive overlap between these patterns (same sequence position, similar structural roles) but also of negative overlap (same sequence position, different structural roles) and of orthogonality (different sequence positions). We suggest that avoidance of negative overlap between patterns of stability determinants is of paramount importance, since such conflicts could rule out any significant intersection between neutral networks. On this score, the two Cro folds rate well, since very few clearly conflicting determinants are suggested by our mutational analysis of P22 Cro and λ Cro, and some of these may even be resolvable. For example, our chameleon design successfully incorporated Tyr at position 45, even though

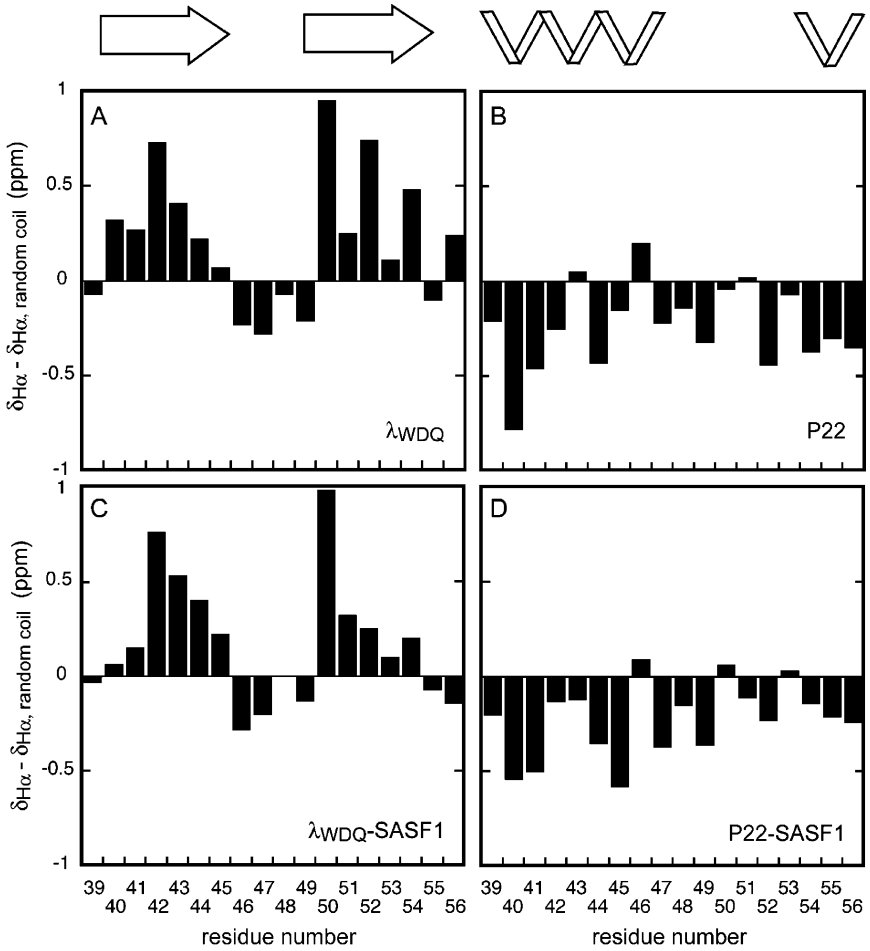

FIGURE 7: Random-coil subtracted H_α chemical shifts for residues 39−56 in (A) λ_{WDQ}, (B) P22, (C) λ_{WDQ}-SASF1, and (D) P22-SASF1. This figure shows that the SASF region adopts the structure of the host sequence into which it is introduced.

mutagenesis studies pointed to a conflict played between the structural roles of Val 45 in P22 Cro and Asn 45 in λ Cro.

Orthogonality between the stability determinants for two folds restricts the possible number of structurally ambivalent sequences but also allows for incorporation of key features encoding the descendant fold through drift without disruption of the ancestral fold. For example, consider the importance of residues in the central strand of λ Cro's β-sheet (residues 40−45; see Figure 4) for the stability of this protein. A successful transition from an all-α and α+β Cro fold would presumably have depended strongly on the sequence of this region. Most of the residues in this region of P22 Cro are not key stability determinants for the all-α fold, being on the outer face of helix 4, or in the case of Leu 42 are important buried residues which are also important in λ Cro (Figure 4E). Neutral sequence drift in this region of an all-α ancestor could have allowed for flexible substitution of residues that facilitated formation of the α+β Cro fold. As an example, Ile 44 in P22 Cro is not structurally important for its all-α fold, but Ile 44 in λ Cro is quite important for its α+β fold. Our SASF or chameleon design (Table 1) reflects this use of orthogonality, incorporating a continuous stretch of wild-type λ Cro residues from position 40 to 44 without significantly destabilizing P22 Cro.

We should note that there may be considerably more flexibility in the way each Cro fold is encoded in the sequence than is apparent from our simple mutagenesis study of P22 Cro and monomeric λ Cro. The scanning point mutagenesis employed here has intrinsic limitations as an approach and offers only the most rough and approximate picture of the stability determinants and sequence space for each of the two Cro folds. First, not all possible residue types at all positions are considered or tested. Second, combinatorial effects are ignored, and conclusions are based on the effect of point mutations within a constant sequence background. Third, P22 Cro and λ Cro, as single representatives of their respective folds, will not cover the full range of detailed structural variation permissible within these topologies, and subtle changes in backbone conformation may alter the favorability of the role played by particular side chains. Fourth, these proteins may not be the most representative possible examples of their respective Cro folds. Our interpretation and comparison of key determinants for different Cro structures are subject to these caveats.

A second important point, not considered in our experiments, is that direct functional selection of residues will influence the likelihood of Cro fold evolution, perhaps as much as selection for folding stability. Although the primary sequence-specific DNA contacts are in the structurally conserved N-terminal half, the C-terminal halves of both α-helical and β-sheet members of the Cro/CI superfamily can also contain direct DNA contacts as well as residues critical for cooperative formation of protein dimers when bound to full DNA sites (*31*, *37*−*39*). Selection for reasons

other than protein stability is thus almost certainly important and could, in principle, either hinder or promote smooth evolutionary transitions between Cro folds. For example, a residue important for function in the ancestral all-α fold could coincide with one important for structure in the descendant α+β fold, and this could help direct Cro evolution down a pathway leading to fold switching. Currently, consideration of the role key C-terminal functional determinants play in Cro structural evolution is limited by the lack of a structure of P22 Cro bound to target DNA.

Our study also omits any consideration of the N-terminal halves of the two Cro sequences. Since the structure of the N-terminal half is largely the same in P22 Cro and λ Cro, one might suppose that few key determinants of evolutionary secondary structure switching are localized to this region. However, the very N-terminus of λ Cro (Figure 1D), lacking in P22 Cro, encodes a β-strand that interacts with the C-terminal β-hairpin (Figure 1C) and could easily influence its formation. The linker between the N- and C-terminal halves also has a different length in the two proteins and could influence the fold, perhaps by causing topological strain in one structure or the other. Finally, one can easily imagine packing interactions between the structurally similar and divergent halves as a factor in determining the overall structure of the protein. All of these possibilities are subjects for future study and, in particular, must be considered if full-length structurally ambivalent Cro-like sequences are to be designed (see below).

The SASFs constructed here represent an advance in the field of protein design. To our knowledge, SASF1 and SASF2, 18−21 residues in length, represent the longest designed chameleon sequences yet constructed, where a chameleon sequence is defined as a polypeptide fragment which undergoes a change in secondary structure when introduced into different tertiary contexts. The first successfully designed chameleons, constructed in the landmark studies of Minor and Kim (*40*), involved sequences 11 residues in length. In addition, the overall sequences of the P22-SASF1 and λ_{WDQ}-SASF1 proteins, which we clearly verified as having different folds, are more than 50% identical (see Table 1), placing them among the most similar pairs of protein sequences known to have different folds (*10*−*12*). In future studies, we hope to design single full-length Cro sequences capable of switching between all-α and α+β folds, to directly address whether overlap exists between the regions of sequence space encoding these structures. The ability to design autonomously folded single sequences of substantial length (∼30 residues) that can directly switch between different domain folds has recently been demonstrated (*14*, *16*).

ACKNOWLEDGMENT

We thank Vahe Bandarian for helpful comments on the manuscript and Neil Jacobsen for technical support with NMR experiments.

SUPPORTING INFORMATION AVAILABLE

A figure showing a comparison of NMR spectra for λ_{WDQ}-SASF1 and λ_{WDQ}-SASF2 (Figure S-1). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Andreeva, A., Howorth, D., Brenner, S. E., Hubbard, T. J., Chothia, C., and Murzin, A. G. (2004) SCOP database in 2004: Refinements integrate structure and sequence family data, *Nucleic Acids Res. 32*, D226−9.
2. Grishin, N. V. (2001) Fold change in evolution of protein structures, *J. Struct. Biol. 134*, 167−85.
3. Lauber, T., Schulz, A., Schweimer, K., Adermann, K., and Marx, U. C. (2003) Homologous proteins with different folds: The three-dimensional structures of domains 1 and 6 of the multiple Kazal-type inhibitor LEKTI, *J. Mol. Biol. 328*, 205−19.
4. Newlove, T., Konieczka, J. H., and Cordes, M. H. (2004) Secondary structure switching in Cro protein evolution, *Structure 12*, 569−81.
5. Babajide, A., Hofacker, I. L., Sippl, M. J., and Stadler, P. F. (1997) Neutral networks in protein space: A computational study based on knowledge-based potentials of mean force, *Folding Des. 2*, 261−9.
6. Bornberg-Bauer, E. (1997) How are model protein structures distributed in sequence space? *Biophys. J. 73*, 2393−403.
7. Govindarajan, S., and Goldstein, R. A. (1997) The foldability landscape of model proteins, *Biopolymers 42*, 427−38.
8. Babajide, A., Farber, R., Hofacker, I. L., Inman, J., Lapedes, A. S., and Stadler, P. F. (2001) Exploring protein sequence space using knowledge-based potentials, *J. Theor. Biol. 212*, 35−46.
9. Cui, Y., Wong, W. H., Bornberg-Bauer, E., and Chan, H. S. (2002) Recombinatoric exploration of novel folded structures: A heteropolymer-based model of protein evolutionary landscapes, *Proc. Natl. Acad. Sci. U.S.A. 99*, 809−14.
10. Dalal, S., and Regan, L. (2000) Understanding the sequence determinants of conformational switching using protein design, *Protein Sci. 9*, 1651−9.
11. Dalal, S., Balasubramanian, S., and Regan, L. (1997) Protein alchemy: Changing β-sheet into α-helix, *Nat. Struct. Biol. 4*, 548−52.
12. Alexander, P. A., Rozak, D. A., Orban, J., and Bryan, P. N. (2005) Directed evolution of highly homologous proteins with different folds by phage display: Implications for the protein folding code, *Biochemistry 44*, 14045−54.
13. Cordes, M. H., Walsh, N. P., McKnight, C. J., and Sauer, R. T. (1999) Evolution of a protein fold in vitro, *Science 284*, 325−8.
14. Cerasoli, E., Sharpe, B. K., and Woolfson, D. N. (2005) ZiCo: A peptide designed to switch folded state upon binding zinc, *J. Am. Chem. Soc. 127*, 15008−9.
15. Cordes, M. H., Burton, R. E., Walsh, N. P., McKnight, C. J., and Sauer, R. T. (2000) An evolutionary bridge to a new protein fold, *Nat. Struct. Biol. 7*, 1129−32.
16. Ambroggio, X. I., and Kuhlman, B. (2006) Computational design of a single amino acid sequence that can switch between two distinct protein folds, *J. Am. Chem. Soc. 128*, 1154−61.
17. Ohlendorf, D. H., Tronrud, D. E., and Matthews, B. W. (1998) Refined structure of Cro repressor protein from bacteriophage λ suggests both flexibility and plasticity, *J. Mol. Biol. 280*, 129−36.
18. LeFevre, K. R., and Cordes, M. H. (2003) Retroevolution of λ Cro toward a stable monomer, *Proc. Natl. Acad. Sci. U.S.A. 100*, 2345−50.
19. Milla, M. E., Brown, B. M., and Sauer, R. T. (1993) P22 Arc repressor: Enhanced expression of unstable mutants by addition of polar C-terminal sequences, *Protein Sci. 2*, 2198−205.
20. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant *m* values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding, *Protein Sci. 4*, 2138−48.
21. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR 6*, 277−93.
22. Johnson, B. A., and Blevins, R. (1994) NMRView. A computer program for the visualization and analysis of NMR data, *J. Biomol. NMR 4*, 603−14.
23. Newlove, T., Atkinson, K. R., Van Dorn, L. O., and Cordes, M. H. (2006) A trade between similar but nonequivalent intrasubunit and intersubunit contacts in Cro dimer evolution, *Biochemistry 45*, 6379−91.
24. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) ^{1}H, ^{13}C

and ^{15}N chemical shift referencing in biomolecular NMR, *J. Biomol. NMR 6*, 135−40.
25. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.
26. Yu, M. H., Weissman, J. S., and Kim, P. S. (1995) Contribution of individual side-chains to the stability of BPTI examined by alanine-scanning mutagenesis, *J. Mol. Biol. 249*, 388−97.
27. Milla, M. E., Brown, B. M., and Sauer, R. T. (1994) Protein stability effects of a complete set of alanine substitutions in Arc repressor, *Nat. Struct. Biol. 1*, 518−23.
28. Mossing, M. C., and Sauer, R. T. (1990) Stable, monomeric variants of λ Cro obtained by insertion of a designed β-hairpin sequence, *Science 250*, 1712−5.
29. Jana, R., Hazbun, T. R., Mollah, A. K., and Mossing, M. C. (1997) A folded monomeric intermediate in the formation of λ Cro dimer-DNA complexes, *J. Mol. Biol. 273*, 402−16.
30. Darling, P. J., Holt, J. M., and Ackers, G. K. (2000) Coupled energetics of λ cro repressor self-assembly and site-specific DNA operator binding I: Analysis of cro dimerization from nanomolar to micromolar concentrations, *Biochemistry 39*, 11500−7.
31. Albright, R. A., and Matthews, B. W. (1998) Crystal structure of λ-Cro bound to a consensus operator at 3.0 Å resolution, *J. Mol. Biol. 280*, 137−51.
32. Viguera, A. R., and Serrano, L. (1999) Stable proline box motif at the N-terminal end of α-helices, *Protein Sci. 8*, 1733−42.
33. Aurora, R., and Rose, G. D. (1998) Helix capping, *Protein Sci. 7*, 21−38.
34. Aurora, R., Srinivasan, R., and Rose, G. D. (1994) Rules for α-helix termination by glycine, *Science 264*, 1126−30.
35. Hutchinson, E. G., and Thornton, J. M. (1994) A revised set of potentials for β-turn formation in proteins, *Protein Sci. 3*, 2207−16.
36. Lattman, E. E., and Rose, G. D. (1993) Protein folding: What is the question? *Proc. Natl. Acad. Sci. U.S.A. 90*, 439−41.
37. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M., and Harrison, S. C. (1988) Recognition of a DNA operator by the repressor of phage 434: A view at high resolution, *Science 242*, 899−907.
38. Beamer, L. J., and Pabo, C. O. (1992) Refined 1.8 Å crystal structure of the λ repressor-operator complex, *J. Mol. Biol. 227*, 177−96.
39. Mondragon, A., and Harrison, S. C. (1991) The phage 434 Cro/OR1 complex at 2.5 Å resolution, *J. Mol. Biol. 219*, 321−34.
40. Minor, D. L., Jr., and Kim, P. S. (1996) Context-dependent secondary structure formation of a designed protein sequence, *Nature 380*, 730−4.
41. Kraulis, P. J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr. 24*, 946−50.
42. Merritt, E. A., and Bacon, D. J. (1997) Raster3D: Photorealistic Molecular Graphics, *Methods Enzymol. 277*, 505−24.

BI060853P