Published in final edited form as:
Proteins. 2013 November ; 81(11): 1988–1996. doi:10.1002/prot.24358.

A role for indels in the evolution of Cro protein folds

Katie L. Stewart, **Michael R. Nelson**, **Karen V. Eaton**, **William J. Anderson**, and **Matthew H. J. Cordes***

[1]Department of Chemistry and Biochemistry, University of Arizona, Tucson, AZ 85721-0088 USA

Abstract

Insertions and deletions in protein sequences, or indels, can disrupt structure and may result in changes in protein folds during evolution or in association with alternative splicing. Pfl 6 and Xfaso 1 are two proteins in the Cro family that share a common ancestor but have different folds. Sequence alignments of the two proteins show two gaps, one at the N terminus, where the sequence of Xfaso 1 is two residues shorter, and one near the center of the sequence, where the sequence of Pfl 6 is five residues shorter. To test the potential importance of indels in Cro protein evolution, we generated hybrid variants of Pfl 6 and Xfaso 1 with indels in one or both regions, chosen according to several plausible sequence alignments. All but one deletion variant completely unfolded both proteins, showing that a longer N-terminal sequence was critical for Pfl 6 folding and a longer central region sequence was critical for Xfaso 1 folding. By contrast, Xfaso 1 tolerated a longer N-terminal sequence with very little destabilization, and Pfl 6 tolerated central region insertions, albeit with substantial effects on thermal stability and some perturbation of the surrounding structure. None of the mutations appeared to convert one stable fold into the other. Based on this two-protein comparison, short insertion and deletion mutations probably played a role in evolutionary fold change in the Cro family, but were also not the only factors.

Keywords

Structural evolution; insertion mutations; deletion mutations; alternative splicing; disruptive mutations

Introduction

Insertion and deletion mutations, or indels, tend to disrupt protein structure[1]. Regions flanking indels show conformational perturbation including destruction of secondary structure, as well as an increase in amino-acid substitution rates [2]. Insertion or deletion of one or a few amino acids within regular secondary structure elements is generally destabilizing, and leads to register shifts that alter side chain interactions or bulges that interrupt backbone topology[3–9]. In sequence alignments of related proteins, indels are an order of magnitude less common than substitutions, and occur mostly in surface loop and turn regions where they are best tolerated structurally[10,11].

*To whom correspondence should be addressed: Name and Address: Matthew Cordes, BSW 439, 1041 E. Lowell St., Tucson, AZ 85721-0088; Phone: 520-626-1175; Fax: 520-626-9204; cordes@email.arizona.edu.

The authors declare no conflicts of interest.

Because of their potential effects on topology, indels may also contribute to the evolution of novel protein architectural elements and new domain folds[12]. On the most disruptive level, some deletions can cause major interruptions in backbone connectivity, forcing the removal and/or replacement of core secondary structure elements. An extreme example is the conversion of a $(\beta\alpha)_8$-barrel domain in the luciferase family to a truncated and topologically distinct barrel, putatively by a deletion of about 90 residues[13,14]. Nondisruptive insertions in surface loops may allow new structural elements and subdomains to bud from existing domains without altering the core fold[15,16]. Loop deletions are one mechanism for evolving or engineering oligomers from monomers through domain swapping[17–20].

Indels may have contributed to the evolution of a novel fold in the bacteriophage Cro proteins. We previously determined the crystal structures of two bacteriophage Cro protein domains, Xfaso 1 and Pfl 6, with 40% sequence identity but different folds (Figure 1)[21]. The all-α fold of Xfaso 1 is common among microbial transcription factors and represents the ancestral form, while the mixed α+β fold of Pfl 6 represents a descendant form in which two to three helices have been replaced with a three-stranded β-sheet[22]. The major structural differences map mostly to the C-terminal portion of the sequence (orange), while the N-terminal halves (silver) share a three-helix subdomain. The extreme N-terminal region also encodes the first strand of the β-sheet in Pfl 6 but is absent in Xfaso 1. The broad distribution of residue identity and similarity across the sequence alignment (Figure 1) suggests that the two proteins have diverged in evolution primarily by accumulation of substitution mutations. However, alignment gaps in two regions also point to a potential role for small indels. One gap occurs near the beginning of the conserved helix 1, possibly coinciding with part of strand 1 of Pfl 6, an element that is not present in Xfaso 1. The other gap occurs in the center of the sequence, near the end of the conserved helix 3 and possibly coinciding with the beginning of helix 4 in Xfaso 1.

Here we assess the importance of these two gaps for Cro structure by generating a series of insertion and/or deletion mutations in Xfaso 1 and Pfl 6 and measuring secondary structure content and thermal stability for each variant. None of the mutations produces an outright change from one stable fold to another, implying that they are insufficient to alter the folds of Xfaso 1 and Pfl 6. Nonetheless, strong mutational effects on thermal folding stability, in particular for the deletions, suggest that a longer N-terminus may be necessary for forming the β-sheet and a longer central region may be important for maintaining the topology of the α-helical fold. Based on this two-protein comparison, both indels and substitutions are probably important for Cro protein fold evolution.

Materials and Methods

Mutagenesis and protein purification

Gene sequences encoding Pfl 6 and Xfaso 1 Cro were previously introduced into pET21b expression vectors, yielding constructs for expression of the wild-type sequences with C-terminal LEHHHHHH tags to enable nickel affinity purification[21]. Insertions and deletions in this study were introduced using the QuikChange method (Stratagene). For Xfaso 1, all indels were introduced into the wild-type sequence background, while for Pfl 6 the indels were constructed an I58D background, with the exception of the VPAER and RARGR

inserts, which were constructed in a wild-type background. Wild-type Pfl 6 and I58D have essentially identical thermal stabilities and circular dichroism spectra at 25 μM protein concentration, but wild-type Pfl 6 dimerizes with a K_d of ~ 1 mM[21] while I58D remains monomeric even at high concentrations due to disruption of a hydrophobic dimer interface[23,24]. The reduced dimerization of I58D facilitates NMR investigations[25,26]. See Table S1 for a complete set of variant sequences used. Proteins were expressed and purified by denaturing Ni-NTA affinity chromatography essentially as described[23] or, in the case of small-scale cultures, by a spin-column version of the same procedure (Qiagen). Purification of Xfaso 1 variants also included 15 mM β-mercaptoethanol in the lysis and wash buffers and 3–5 mM β-mercaptoethanol in the elution buffer to prevent disulfide bond formation. For uniform ^{15}N-labelled samples, proteins were expressed in M9T minimal medium containing 0.8 g/L $^{15}NH_4Cl$ as the sole nitrogen source, and purified in the same manner as unlabelled proteins. Purified proteins were refolded by dialysis into SB250 buffer [50 mM Tris (pH 7.5), 250 mM KCl, and 0.2 mM EDTA], and dialysates were centrifuged to remove precipitates prior to measurement of soluble protein concentrations. In the case of Xfaso 1 variants, 1 mM DTT was included to prevent disulfide bond formation. Protein concentrations were estimated by ultraviolet absorbance at 280 nm using extinction coefficients of 3119 cm^{-1} for Pfl 6 and 5559 cm^{-1}. Extinction coefficients for Pfl 6 variants lacking one of the two wild-type tyrosine residues were estimated as half the wild-type value. Extinction coefficients for Xfaso 1 variants containing an additional tyrosine were estimated as 6756 cm^{-1}.

Circular dichroism spectroscopy

Circular dichroism spectra and thermal denaturation curves were obtained on an OLIS DSM-20 CD spectrophotometer. Wavelength scans were obtained at 20 °C at a protein concentration of 25 μM in a 1 mm pathlength cylindrical cell, from 260 to 205 nm in 1 nm steps with an integration time of 5–30 s. Signals were averaged from 3–5 scans, and spectra were corrected for buffer baseline signals. Thermal denaturation curves were obtained at a protein concentration of 25 μM in a 2 mm pathlength cylindrical cell, from 20–80 °C in 2 °C steps, with 2 min equilibration time and 25–55 s signal integration time for each temperature point. T_m values were obtained by fitting to the following relationship[27]:

$$\Delta G_u = \Delta H_u\,(1 - T/T_m) + \Delta C_p\,[T - T_m - T * \ln(T/T_m)]$$

The free energy of unfolding ΔG_u relates directly to the fraction of molecules in the unfolded state, and this fraction in turn relates to the position of the measured ellipticity value relative to the unfolded and folded baselines. Baseline slopes and intercepts were allowed to vary in fits. The heat capacity of unfolding (ΔC_p) was fixed at 840 cal mol^{-1} K^{-1} based on an estimate of 14 cal mol^{-1} K^{-1} per residue for a 60-residue folded region[28].

NMR spectroscopy

^{15}N-^{1}H NMR correlation spectra were recorded at 293 K on a Varian Inova 600 MHz spectrometer equipped with a triple-resonance cryogenic probe. Spectra were processed using NMRPipe/NMRDraw [29] and analyzed using Sparky (T.D. Goddard and D.G. Kneller,

SPARKY 3, University of California, San Francisco). Pfl 6 I58D backbone resonance assignments were originally obtained by strip plot analysis of 3D NOESY-HSQC and 3D TOCSY-HSQC experiments in samples containing 2 mM protein in 50 mM phosphate (pH 6.0–6.1) at 293 K. Most of these assignments were readily transferred by inspection of spectral overlays to samples of Pfl 6 I58D containing 1.0–1.5 M TMAO or TMAO-d^9, thereby allowing analysis of effects of the VPAER and GGGGG insertions on specific backbone resonances (see Fig. 5 and Fig. S1).

Results

Double indel variants

We aligned the Xfaso 1 and Pfl 6 sequences in three different ways with respect to the N-terminal gap, and in two different ways with respect to the central region (Figure 1). Based on the alignments, we constructed twelve hybrid variants each containing insertion and deletion mutations in both the N-terminal and central regions (refer to Table S1 for full sequences of indel variants used in this study). The Pfl 6 variants were constructed in an I58D background[25,26], which reduces solution dimerization but maintains the α+β fold and stability of individual subunits[21,24]. We purified each variant and characterized its structure and stability using far ultraviolet circular dichroism (CD) wavelength scans and thermal denaturation curves (Figure 2). Wavelength scans of the Pfl 6 (Figure 2A) and Xfaso 1 variants (Figure 2C) at 20 °C show much weaker ellipticity at 222 nm compared to the parent proteins, as well as a much stronger signal at 205 nm than at 222 nm. Thermal melts of all variants (Figures 2B and 2D) have essentially no temperature dependence of dichroism at 222 nm. These observations combine to indicate that all twelve hybrid Pfl 6 and Xfaso 1 variants are completely unfolded at 20 °C. Loss of folding stability shows that insertions and deletions at the N-terminus and central linker region are important structural determinants; however, they are insufficient by themselves to convert Xfaso 1 to a stable partly β-sheet protein or Pfl 6 to a stable all α-helical protein. Other features of the sequences of these two proteins must also be important for specifying and stabilizing their folds.

N-terminal indels

We then studied indels in the N-terminal and central regions separately. Two-residue N-terminal deletions in Pfl 6 (Figures 3A and 3B) severely impact folding stability. Neither PLdel nor KYdel give any evidence of significant population of a folded state. The far ultraviolet CD spectrum of KKdel has properties intermediate between native wild-type Pfl 6 and the putatively unfolded PLdel and KYdel. Its thermal melt does show some loss of signal with increasing temperature, indicative of a partial thermal unfolding transition. Assuming that the spectrum of fully native KKdel resembles that of the wild type, the level of signal at 20 °C suggests that the value of T_m is near ambient temperature, compared to T_m for the wild-type protein of 64 °C. In sum, the Pfl 6 structure tolerates N-terminal deletions poorly, with some dependence on the specific location: even a two-residue deletion at the very N-terminus strongly destabilizes Pfl 6, and more internal deletions lead to complete unfolding.

The structure of Xfaso 1 tolerates N-terminal insertions (Figures 3C and 3D) rather well. KKins has a spectrum and melt almost identical to those of the wild type, suggesting no significant mutational effect on stability or structure. PLins has a slightly weaker spectrum and a thermal melt with a T_m of 45 °C, compared to 51 °C for the wild type. Thus, this variant appears mostly folded but slightly destabilized relative to the wild type. KYins appears completely unfolded based on an interpretation similar to that used for the Pfl 6 variants (see above). In sum, Xfaso 1 tolerates some two-residue insertion mutations in this region quite well, but effects on stability become progressively more severe for mutations more internal to the sequence.

The apparent tolerance of Xfaso 1 for the internal PL insertion at residues 5–6 suggests that the structure can accommodate both extra residues at the N-terminus as well as displacement of existing residues further toward the N-terminus. To test this idea, we generated a variant, PLins NK2/A3K, in which the first four residues (MNAI) were completely replaced by the first six residues of Pfl 6 (MKKIPL). This variant showed properties essentially identical properties to those of PLins, suggesting that this region of Xfaso 1 is in fact quite tolerant of wide sequence variation at the N-terminus, at least in the region of residues 1–4.

Central region indels

Insertion of either the VPAER or RARGR sequence from Xfaso 1 into the aligned positions in Pfl 6 (Figures 4A and 4B) results in apparent strong destabilization in both cases, but not complete unfolding. Both the VPAERins and RARGRins variants display some indication of a thermal unfolding transition, though this is less apparent for RARGRins. Both also show ellipticities at 222 nm that are intermediate between those of wild type Pfl 6 and unfolded variants. Similar to the KK deletion at the N-terminus, these are probably unstable variants, with the denaturation midpoint of VPAERins close to ambient temperatures, and that of RARGRins somewhat lower. Deletion of the VPAER or RARGR sequence in Xfaso 1 (Figures 4C and 4D) results in complete unfolding with behavior similar to that observed for some of the Pfl 6 N-terminal deletion variants. In sum, the structure of Pfl 6 exhibits some low-to-moderate tolerance for insertion mutations in the central region, while deletion mutations in Xfaso 1 lead to catastrophic destabilization.

All insertion and deletion variants of Xfaso 1 characterized in this study show either complete unfolding, or clear retention of greater helical secondary structure than that seen for Pfl 6. Consistent with the conclusions from the double indel variants, none of the mutations alone can convert the native fold of Xfaso 1 from a stable α-helical form to a stable α+β form. Similarly, none of the Pfl 6 variants show evidence for increased native helicity expected from conversion to the all α-helical fold. Pfl 6 variants such as VPAERins and KKdel in fact show reduced ellipticity at 222 nm at 20 °C despite displaying unfolding transitions. As noted above, this reduced baseline ellipticity might result from a significant population of unfolded protein molecules even at low temperature, but it might also have contributions from a reduction in native helical content caused by structural disruption of the folded state.

Further structural characterization of Pfl 6 VPAERins

We used NMR to characterize structural effects of the Pfl 6 VPAER insertion in more detail. Initial spectra of uniform ^{15}N-labelled Pfl 6-VPAERins-I58D variant in phosphate buffers showed extremely poor spectral quality and dispersion, possibly due to a population of unfolded molecules in exchange with the native state. Addition of trimethylamine oxide (TMAO), an osmolyte that specifically stabilizes the native states of proteins,[30–32] to a concentration of 1.5 M improved spectral quality (Figure 5), though it also caused some sample precipitation. This spectrum was of sufficient quality for a limited comparison with that of Pfl 6-I58D, for which we had a full set of backbone resonance assignments (Figure 5).

Numerous well-separated resonances within different strands of the β-sheet (strand 1: residues 2–5, strand 2: 40–44, strand 3: 50–56) showed close correspondences between VPAERins-I58D and the reference I58D variant, indicating retention of the β-sheet framework. Interestingly, however, the spectrum of VPAERins-158D shows some severely weakened/absent/shifted resonances assignable to the helix-turn-helix (16–36) region, along with some probable resonance doubling within this motif (note particularly the side chain NH_2 group for residue 16 in Figure 5). A likely source of the doubling effects is proline *cis-trans* isomerization within the inserted sequence, coupled to changes in conformation elsewhere. The spectrum of a GGGGGins-I58D variant shows some similar resonance perturbations (Figure S1) but better spectral quality at lower TMAO concentration (1 M), no clear resonance doubling, and stronger resonances in the helix-turn-helix. This suggests that five-residue insertions perturb the structure of Pfl 6 regardless of sequence and conformation, but not in exactly the same way, and more flexible sequences may be less disruptive. Changes in the structure and dynamics of the helix-turn-helix region may be partly responsible for the reduced low temperature helicity seen in the circular dichroism spectra of VPAERins, though low folding stability probably also contributes (Figure 4). In sum, insertion of a five-residue sequence from Xfaso 1 into the central region of Pfl 6 does not convert the C-terminus from β-sheet to α-helix, but it does cause global destabilization and structural perturbation, as one might expect.

Discussion

Xfaso 1 and Pfl 6 Cro have different folds. Alignments of their sequences show ~40% identity and two clear alignment gaps, one at the N-terminus and one in the middle of the sequence. The gaps and their location suggest that differences in sequence length could be important for specifying the two different folds, and that indels could have played a role in evolutionary conversion of the all-α fold to the α+β fold. Here we have shown that mutations that ablate both alignment gaps also abolish folding in both proteins. Thus, differences in sequence length at the N-terminus and central region are clearly important for their differences in fold, but cannot be solely responsible for them; substitutions must also be required to convert the fold of either Xfaso 1 or Pfl 6 to an alternate stable form. As just one example, the polar residue R47 on the surface of in helix 4 of Xfaso 1 aligns to an interior-facing position in strand 2 of Pfl 6, and a polar residue at this position may hinder formation of the α+β fold. In the mechanism of Cro fold evolution, both indels and

substitutions were probably important, though their relative importance and the order of mutational events remains unclear.

Deletion mutations are generally more disruptive to structure than insertions[1], and those studied here are no exception. While some of the insertions in Xfaso 1 and Pfl 6 are structurally tolerated, all of the deletion mutations, except KKdel in Pfl 6, completely destroy folding. These effects on stability suggest that the helical fold of Xfaso 1 requires a relatively long central region linking the N- and C-terminal subdomains, while the α+β fold of Pfl 6 requires a relatively long N-terminus. One implication is that a stable evolutionary bridge sequence between the two folds (if it existed) would most likely have both a long N-terminus and a long central region. Such an ambivalent sequence could then be converted completely to one fold or the other by a deletion in one of the two places.

Deletion of N-terminal residues of Pfl 6 probably destabilizes the α+β fold by eliminating part of strand 1 of the three-stranded β-sheet, though the positional dependence of the effect shows that both backbone and side chain effects are important. Thus, one factor in the inability of Xfaso 1 to adopt a β-sheet fold is that it could not easily form the first strand of the sheet to interact with the two strands in the C-terminal region. No N-terminal strand exists in the helical fold, allowing Xfaso 1 to have a shorter N-terminus. However, Xfaso 1 can also be extended at the N-terminus without much destabilization. We speculate that non-disruptive N-terminal extension of a helical Cro protein could have been a prerequisite for evolutionary structure switching.

Deletion of the central region of Xfaso 1 most likely unfolds this protein by generating a topological discontinuity in which the end of helix 3 and the beginning of helix 4 cannot easily be connected (see Figure 1). The Pfl 6 structure, on the other hand, readily accommodates a shorter sequence in the region connecting helix 3 to strand 2, and in fact favors it. If one views the five-residue linker as part of the helical C-terminal subdomain of Xfaso 1 (the VPAER sequence includes the beginning of helix 4), one possible explanation of these different preferences is that the helical C-terminal conformation requires a longer sequence to form correctly than does the β-sheet conformation. Helical secondary structure elements are intrinsically coiled and cover a distance of about 1.5 Å distance per residue along the helical axis, much less than the value of about 3.5 Å per residue for extended β-strand elements. In effect, deletions may tend to unravel helices and favor strand formation to maintain topological continuity. Grishin has proposed that many substitutions of structural elements in folds may result from the topological effects of indels[12]. Of course, in the case of our Xfaso 1 deletions, unfolding occurs instead of fold switching, presumably because other features of the Xfaso 1 sequence prevent stable formation of the alternate β-sheet fold. But because of its strong effect, a deletion in the central region could easily have been the proximate cause for fold switching during Cro evolution.

A limitation of this work is that mutagenesis studies on Xfaso 1 and Pfl 6, the most similar pair of Cro sequences known to have the ancestral and descendant folds, respectively, cannot fully represent the different sequence determinants for these two folds or the possible evolutionary pathways between them. The conclusions of this study are limited to whether simple insertions and deletions in these two representatives could suffice to explain major

structural changes in Cro fold evolution. The answer is no, but as far as the present comparison goes such indels are likely to be a necessary component of the mechanism. A more complete understanding may emerge from a phylogenetic analysis of the family, currently in progress.

Indels have effects on protein structure beyond those produced by substitutions alone. The unique properties of indels may be important not only in structural evolution but also in protein engineering[1] and in modulation of protein structure and function by alternative splicing in eukaryotes[33,34]. Many changes in sequence length will, if disruptive, strongly destabilize a protein structure and potentially inactivate the protein, but the structural plasticity of proteins also allows accommodation of disruptive indels within folded structures in varied and surprising ways[35]. These include the possibility of changes in fold but also, for example, extensive translocations of sequence that preserve folding topology[36]. The effects of alternative splicing on protein structure remain to be thoroughly explored, and many splicing events probably have either minor effects on structure or lead to global unfolding. But in some cases two alternatively spliced isoforms fold with significant differences in folding topology and function[37]. This phenomenon may prove to be widespread[33]. Whether produced by alternative splicing, evolutionary mutations or engineering, indels have a key role to play in the generation of novel protein architecture.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institute for General Medical Sciences at the National Institutes of Health [grant number R01 GM066806 to M.H.J.C.].

References

1. Shortle D, Sondek J. The emerging role of insertions and deletions in protein engineering. Curr Opin Biotechnol. 1995; 6(4):387–393. [PubMed: 7579648]
2. Zhang Z, Huang J, Wang Z, Wang L, Gao P. Impact of indels on the flanking regions in structural domains. Mol Biol Evol. 2011; 28(1):291–301. [PubMed: 20671041]
3. Heinz DW, Baase WA, Dahlquist FW, Matthews BW. How amino-acid insertions are allowed in an alpha-helix of T4 lysozyme. Nature. 1993; 361(6412):561–564. [PubMed: 8429913]
4. Heinz DW, Baase WA, Zhang XJ, Blaber M, Dahlquist FW, Matthews BW. Accommodation of amino acid insertions in an alpha-helix of T4 lysozyme. Structural and thermodynamic analysis. J Mol Biol. 1994; 236(3):869–886. [PubMed: 8114100]
5. Keefe LJ, Quirk S, Gittis A, Sondek J, Lattman EE. Accommodation of insertion mutations on the surface and in the interior of staphylococcal nuclease. Protein Sci. 1994; 3(3):391–401. [PubMed: 8019410]
6. Keefe LJ, Sondek J, Shortle D, Lattman EE. The alpha aneurism: a structural motif revealed in an insertion mutant of staphylococcal nuclease. Proc Natl Acad Sci U S A. 1993; 90(8):3275–3279. [PubMed: 8475069]
7. Sondek J, Shortle D. Accommodation of single amino acid insertions by the native state of staphylococcal nuclease. Proteins. 1990; 7(4):299–305. [PubMed: 2381904]
8. Sondek J, Shortle D. Structural and energetic differences between insertions and substitutions in staphylococcal nuclease. Proteins. 1992; 13(2):132–140. [PubMed: 1620695]

9. Vetter IR, Baase WA, Heinz DW, Xiong JP, Snow S, Matthews BW. Protein structural plasticity exemplified by insertion and deletion mutants in T4 lysozyme. Protein Sci. 1996; 5(12):2399–2415. [PubMed: 8976549]
10. Benner SA, Cohen MA, Gonnet GH. Empirical and structural models for insertions and deletions in the divergent evolution of proteins. J Mol Biol. 1993; 229(4):1065–1082. [PubMed: 8445636]
11. Pascarella S, Argos P. Analysis of insertions/deletions in protein structures. J Mol Biol. 1992; 224(2):461–471. [PubMed: 1560462]
12. Grishin NV. Fold change in evolution of protein structures. J Struct Biol. 2001; 134(2–3):167–185. [PubMed: 11551177]
13. Fisher AJ, Raushel FM, Baldwin TO, Rayment I. Three-dimensional structure of bacterial luciferase from Vibrio harveyi at 2. 4 A resolution. Biochemistry. 1995; 34(20):6581–6586. [PubMed: 7756289]
14. Moore SA, James MN, O'Kane DJ, Lee J. Crystal structure of a flavoprotein related to the subunits of bacterial luciferase. EMBO J. 1993; 12(5):1767–1774. [PubMed: 8491169]
15. Blouin C, Butt D, Roger AJ. Rapid evolution in conformational space: a study of loop regions in a ubiquitous GTP binding domain. Protein Sci. 2004; 13(3):608–616. [PubMed: 14978301]
16. Jiang H, Blouin C. Insertions and the emergence of novel protein structure: a structure-based phylogenetic study of insertions. BMC Bioinformatics. 2007; 8:444. [PubMed: 18005425]
17. Albright RA, Mossing MC, Matthews BW. High-resolution structure of an engineered Cro monomer shows changes in conformation relative to the native dimer. Biochemistry. 1996; 35(3): 735–742. [PubMed: 8547253]
18. Bennett MJ, Schlunegger MP, Eisenberg D. 3D domain swapping: a mechanism for oligomer assembly. Protein Sci. 1995; 4(12):2455–2468. [PubMed: 8580836]
19. Green SM, Gittis AG, Meeker AK, Lattman EE. One-step evolution of a dimer from a monomeric protein. Nat Struct Biol. 1995; 2(9):746–751. [PubMed: 7552745]
20. Schlunegger MP, Bennett MJ, Eisenberg D. Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. Adv Protein Chem. 1997; 50:61–122. [PubMed: 9338079]
21. Roessler CG, Hall BM, Anderson WJ, Ingram WM, Roberts SA, Montfort WR, Cordes MH. Transitive homology-guided structural studies lead to discovery of Cro proteins with 40% sequence identity but different folds. Proc Natl Acad Sci U S A. 2008; 105(7):2343–2348. [PubMed: 18227506]
22. Newlove T, Konieczka JH, Cordes MH. Secondary structure switching in Cro protein evolution. Structure. 2004; 12(4):569–581. [PubMed: 15062080]
23. LeFevre KR, Cordes MH. Retroevolution of lambda Cro toward a stable monomer. Proc Natl Acad Sci U S A. 2003; 100(5):2345–2350. [PubMed: 12598646]
24. Newlove T, Atkinson KR, Van Dorn LO, Cordes MH. A trade between similar but nonequivalent intrasubunit and intersubunit contacts in Cro dimer evolution. Biochemistry. 2006; 45(20):6379–6391. [PubMed: 16700549]
25. Bouvignies G, Korzhnev DM, Neudecker P, Hansen DF, Cordes MH, Kay LE. A simple method for measuring signs of (1)H (N) chemical shift differences between ground and excited protein states. J Biomol NMR. 2010; 47(2):135–141. [PubMed: 20428928]
26. Bouvignies G, Vallurupalli P, Cordes MH, Hansen DF, Kay LE. Measuring 1HN temperature coefficients in invisible protein states by relaxation dispersion NMR spectroscopy. J Biomol NMR. 2011; 50(1):13–18. [PubMed: 21424227]
27. Becktel WJ, Schellman JA. Protein stability curves. Biopolymers. 1987; 26(11):1859–1877. [PubMed: 3689874]
28. Myers JK, Pace CN, Scholtz JM. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. Protein Sci. 1995; 4(10):2138–2148. [PubMed: 8535251]
29. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 1995; 6(3):277–293. [PubMed: 8520220]
30. Baskakov I, Bolen DW. Forcing thermodynamically unfolded proteins to fold. J Biol Chem. 1998; 273(9):4831–4834. [PubMed: 9478922]

31. Mello CC, Barrick D. Measuring the stability of partly folded proteins using TMAO. Protein Sci. 2003; 12(7):1522–1529. [PubMed: 12824497]
32. Wang A, Bolen DW. A naturally occurring protective system in urea-rich cells: mechanism of osmolyte protection of proteins against urea denaturation. Biochemistry. 1997; 36(30):9101–9108. [PubMed: 9230042]
33. Birzele F, Csaba G, Zimmer R. Alternative splicing and protein structure evolution. Nucleic Acids Res. 2008; 36(2):550–558. [PubMed: 18055499]
34. Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, Cortese MS, Sickmeier M, LeGall T, Obradovic Z, Dunker AK. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. Proc Natl Acad Sci U S A. 2006; 103(22):8390–8395. [PubMed: 16717195]
35. Schellenberg MJ, Ritchie DB, Wu T, Markin CJ, Spyracopoulos L, MacMillan AM. Context-dependent remodeling of structure in two large protein fragments. J Mol Biol. 2010; 402(4):720–730. [PubMed: 20713060]
36. Sagermann M, Baase WA, Matthews BW. Sequential reorganization of beta-sheet topology by insertion of a single strand. Protein Sci. 2006; 15(5):1085–1092. [PubMed: 16597830]
37. Garcia J, Gerber SH, Sugita S, Sudhof TC, Rizo J. A conformational switch in the Piccolo C2A domain regulated by alternative splicing. Nat Struct Mol Biol. 2004; 11(1):45–53. [PubMed: 14718922]

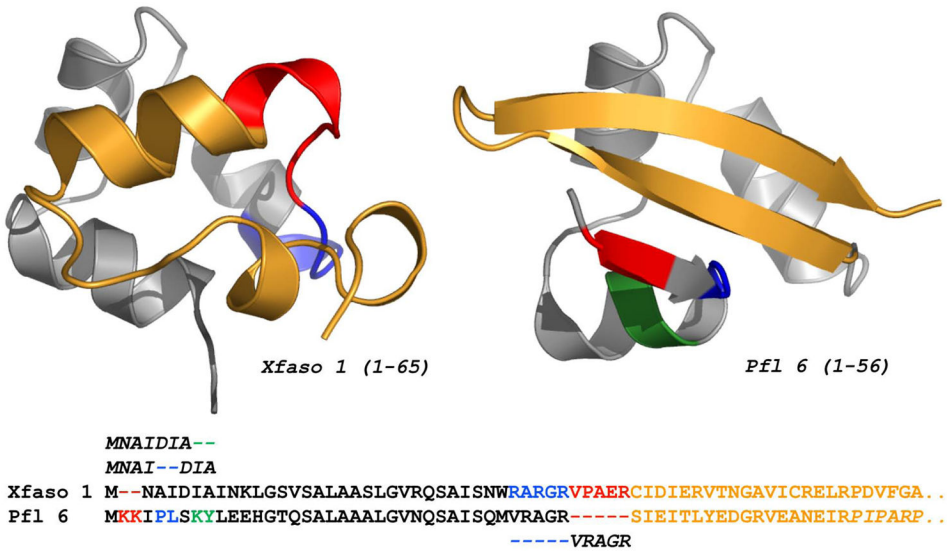

Figure 1.

Structures and sequence alignment of Xfaso 1 and Pfl 6, two Cro proteins with ~40% sequence identity but different folds. Both proteins share a conserved N-terminal subdomain (gray/black) with three helices and a divergent C-terminal subdomain (orange) with either helical (Xfaso 1) or strand (Pfl 6) secondary structure. Residues 57–62 of Pfl 6 (PIPARP, in italics) are not shown in the structure. The two alignment gaps map to near the beginning or end of the conserved subdomain. Xfaso 1 has a five-residue insertion relative to Pfl 6 between the two subdomains, either at the beginning of helix 4 in the C-terminal subdomain (red) or at the end of the conserved third helix (blue), depending on how the sequences are aligned (alternate alignment indicated by italics). Similarly, Pfl 6 has a two residue N-terminal extension that, depending on alignment, localizes either to a short β-strand (red) or to the first turn of helix 1 (blue or green).

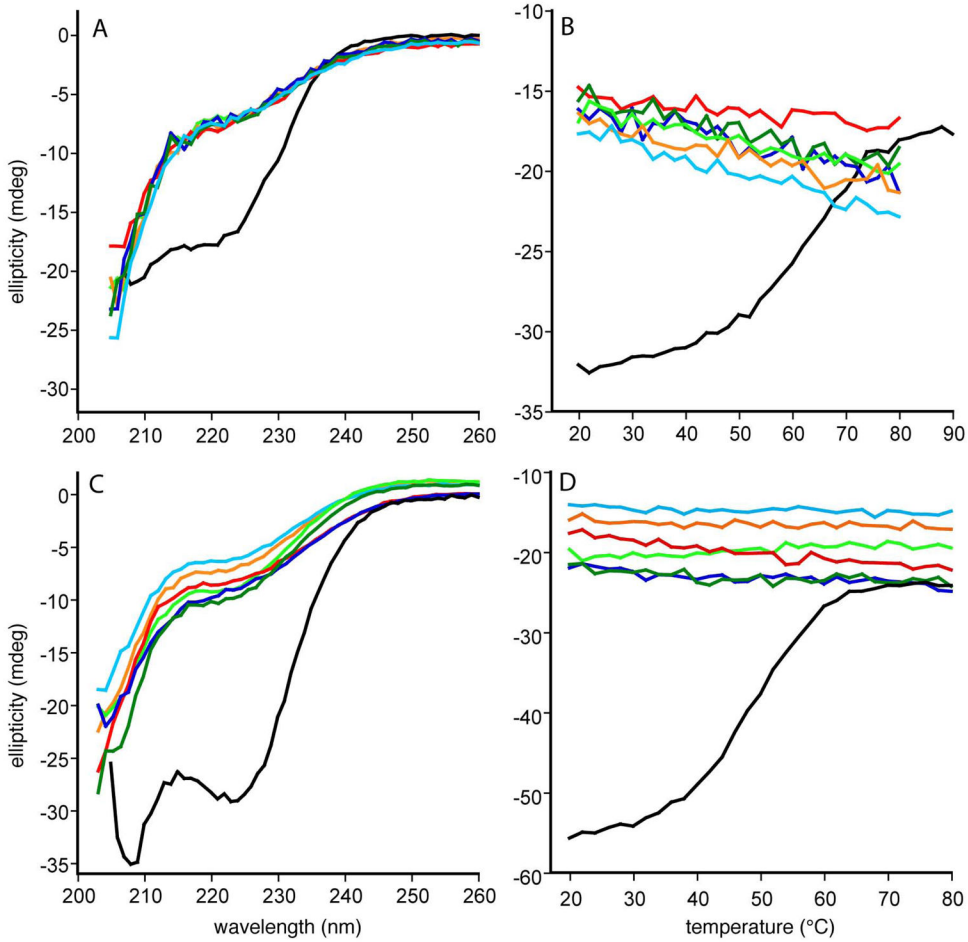

Figure 2.

Double indel variants of Pfl 6 and Xfaso 1 are unfolded: A) far ultraviolet circular dichroism spectra and B) melts of Pfl 6 variants; C) far ultraviolet circular dichroism spectra and D) melts of Xfaso 1 variants. Each panel shows data for seven variants: wild-type (black), KK/VPAER (orange), PL/VPAER (cyan), KY/VPAER (light green), KK/RARGR (red), PL/RARGR (blue) and KY/RARGR (green). In the shorthand notation for variants, the first two residues represent an N-terminal deletion in Pfl 6 or insertion in Xfaso 1, according to the alignments shown in Figure 1; similarly, the last five residues represent a central region insertion in Pfl 6 or deletion in Xfaso 1.

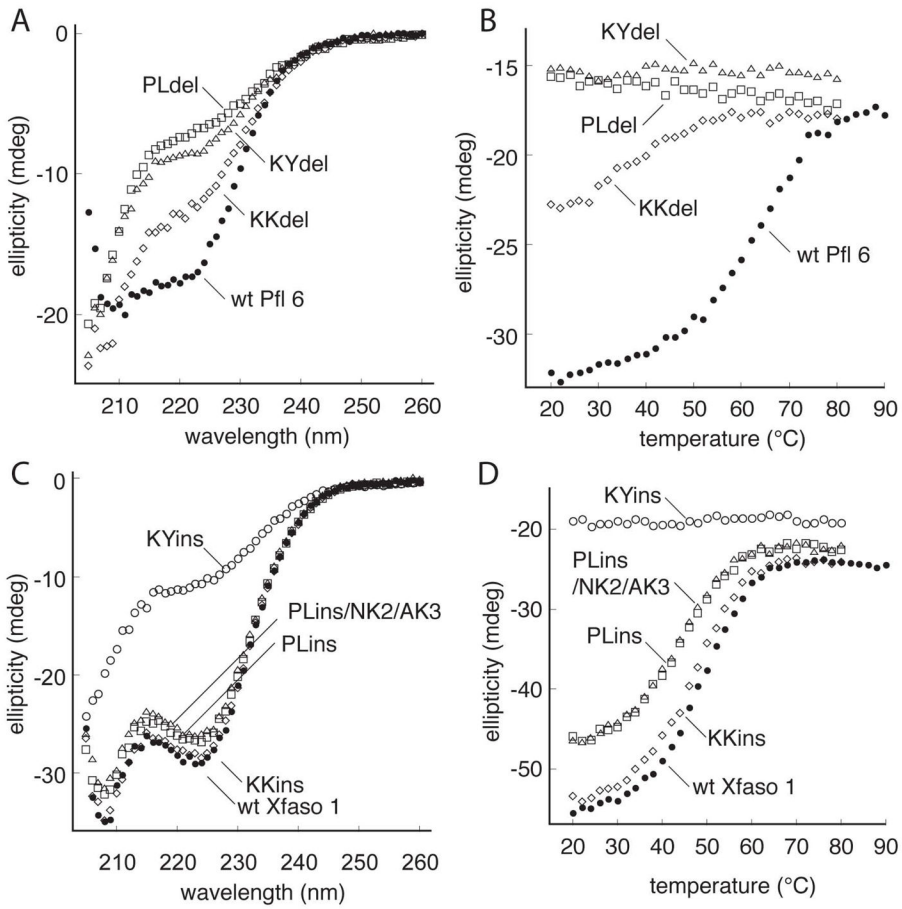

Figure 3.

N-terminal indel variants: A) far ultraviolet circular dichroism spectra and B) melts of Pfl 6 variants; C) far ultraviolet circular dichroism spectra and D) melts of Xfaso 1 variants. Variants correspond to either deletion (Pfl 6) or insertion (Xfaso 1) of two residues according to the alignments shown in Figure 1. PLins/NK2/AK3 corresponds to replacing residues 1–4 of Xfaso 1 (MNAI) with residues 1–6 (MKKIPL) of Pfl 6.

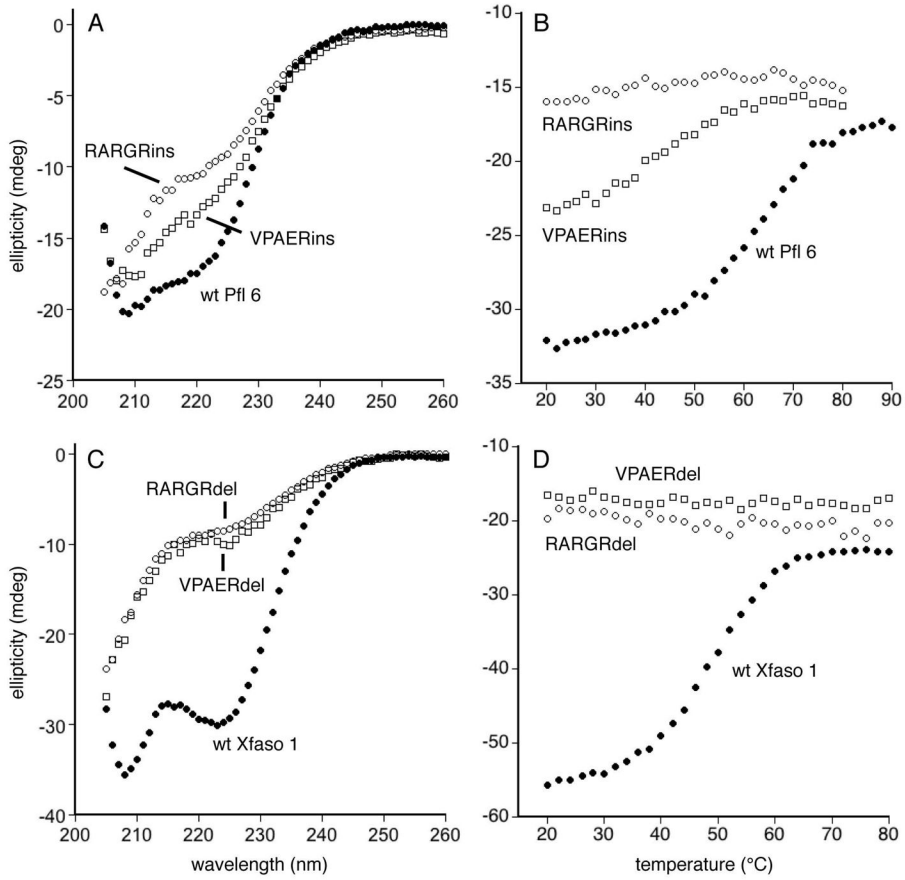

Figure 4.

Central region indel variants: A) far ultraviolet circular dichroism spectra and B) melts of Pfl 6 variants; C) far ultraviolet circular dichroism spectra and D) melts of Xfaso 1 variants. Variants correspond to either insertion (Pfl 6) or deletion (Xfaso 1) of five residues according to the alignments shown in Figure 1.

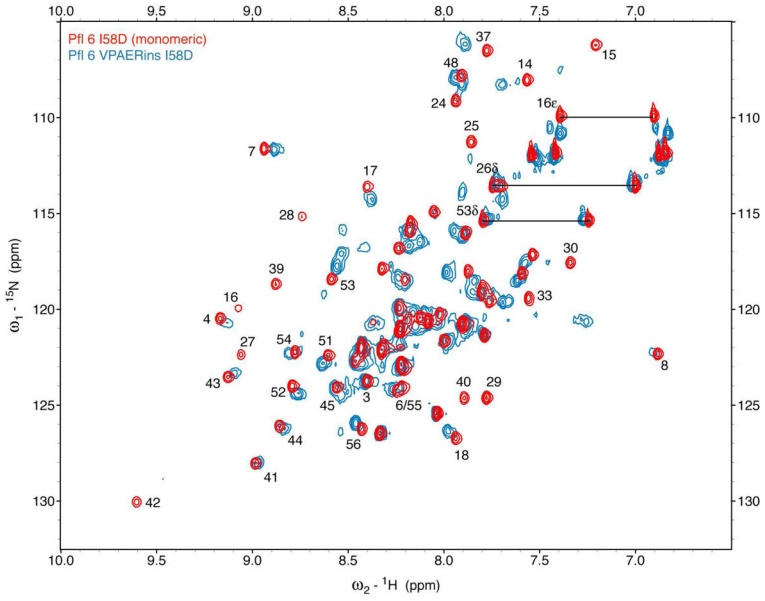

Figure 5.

Structural effects of five-residue central region insertion VPAER on Pfl 6: 2D ^{15}N-^{1}H correlation spectra of Pfl 6 I58D and Pfl 6 I58D VPAERins, in 50 mM sodium phosphate (pH 7) containing 1.5 M TMAO. Nominal protein concentrations are ~1 mM but addition of TMAO caused some precipitation, particularly for the VPAERins variant. Peak assignments are shown for relevant and/or well-resolved resonances of Pfl 6 I58D.