Minireview

α-Helical Protein Assembly Motifs*

Wayne D. Kohn†, Colin T. Mant§, and Robert S. Hodges$$

From the Department of Biochemistry, the MRC Group in Protein Structure and Function and the Protein Engineering Network of Centres of Excellence, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

This review will focus on α-helical protein assembly motifs where the α-helix is the major element of secondary structure involved in the folding and stability of the structure and may also be involved in function by binding to receptor molecules. Apart from the three types of α-helical motifs discussed, i.e. those motifs that form autonomously folded protein domains; those motifs that only form a stable folded domain when dimerized; and a motif that requires other structural elements to contribute to the hydrophobic core to stabilize a folded domain, we will present examples of more complex protein assemblies that have combined two different motifs to form a functional molecule.

Two-, Three-, and Four-stranded Coiled-coil Motifs

α-Helical coiled-coils represent what is probably the most widespread assembly motif found in proteins. A coiled-coil model was first proposed by Crick in 1953 (1) and is comprised of two, three, or four right-handed amphipathic α-helices, which wrap around each other in a left-handed supercoil with a crossing angle of approximately 20° between helices (Fig. 1, F and G) such that their hydrophobic surfaces are in continuous contact to form, respectively, dimeric, trimeric, or tetrameric coiled-coils. The formation of a coiled-coil is dependent primarily on the presence of heptad repeat sequences of a form denoted [abcdef]n, where positions a and d are characteristically occupied by hydrophobic residues (i.e. a 3–4 or 4–3 hydrophobic repeat) (2–4). The hydrophobic a and d residues form the core of the coiled-coil, while the e and g positions flank the hydrophobic interface, packing against the residues of the hydrophobic core, and may also participate in interhelical g–e contacts (5–7). Two-stranded coiled-coils have traditionally been recognized as a dimerization unit in fibrous proteins such as tropomyosin (284 residues) and myosin (approximately 1100 residues) as well as the longest coiled-coil found so far, NuMA (1485 residues) (8). Subsequently, the motif has been discovered in a wide variety of proteins (2, 9–11). Several NMR and crystallographic studies have made clear the inappropriateness of the “zipper” description first applied to the so-called basic leucine zipper (bZIP) class of eukaryotic transcription factors, which are, in fact, traditional coiled-coils. Structures have also been determined for the GCN4 and Fos/Jun factors bound to their DNA recognition sequences (6, 7, 12). Dimeric coiled-coils have also been found in many other contexts. For example, a 39-residue coiled-coil mediates the dimerization of cyclic GMP-dependent protein kinase (13). Other proteins that oligomerize through formation of dimeric coiled-coils include the β dimer of the G signal transduction protein complex (14) and transcription factors of the basic helix-loop-helix leucine zipper (bHLH-ZIP) (15) and homeodomain-ZIP (16) classes, as well as transcription factors GAL4 (17) and PPR1 (18). In the bZIP, bHLH-ZIP, and homeodomain-ZIP proteins, the coiled-coil is linked directly to the DNA binding motif (Fig. 1), and exact spacing is critical for activity. In contrast, the GAL4 and PPR1 yeast transcription factors contain an extended 9-residue linker between the DNA binding motif and short coiled-coils of 14 and 19 residues, respectively (Fig. 1). Although the majority of two-stranded coiled-coils have parallel strands, a number of coiled-coils with antiparallel strands have been observed. Such coiled-coils may be intrachain, where the coiled-coil is formed by two helices joined by a turn, or they can be interchain interactions between separate polypeptide chains. Due to such antiparallel alignment, the packing of residues at the dimer interface is different than in parallel coiled-coils (19). It has been shown that the interhelical ionic interactions involving these residues can affect coiled-coil orientation (19). Examples of interchain dimerization to form antiparallel coiled-coils are observed in the crystal structures of the replication terminator protein of B. subtilis (20) and pilin protein of N. gonorrhoeae (21). Examples of the more common intrachain antiparallel coiled-coils have been found in several enzymes, including bacterial seryl-tRNA synthetase (22), the GREA transcript cleavage factor of E. coli (23), and the T-cell protein tyrosine kinase ZAP-70 (24). The simplicity of the dimeric coiled-coil structure makes it an ideal system to use in understanding the fundamentals of protein folding and stability and in testing the principles of de novo design for a wide range of medical applications (4).

α-Fibrous proteins, containing the peptidic repeat characteristic of coiled-coils, which exhibit trimeric rather than dimeric oligomerization, include laminin, tenasin, fibrinogen, and macrophage scavenger receptor protein (25, 26). Interestingly, extracellular α-fibrous proteins are three-stranded compared with the two-stranded α-fibrous proteins found intracellularly. Other proteins containing trimeric coiled-coil domains include influenza hemagglutinin (27), heat shock transcription factors (28), spectrin (29), α-actinin (30), dystrophin (31), and mannose-binding proteins (32). A clear picture of the motif has emerged from the crystal structures of, for instance, a soluble trimeric fragment of influenza hemagglutinin (27) (Fig. 1C), synthetic GCN4 leucine zipper mutants (33, 34), and a trimeric fragment of C-type mannose-binding proteins (32).

There is a relative lack of reports concerning stability and folding of four-chain coiled-coils compared with dimeric coiled-coils (35). However, some tetrameric coiled-coils have been characterized (36, 37). X-ray crystallography of GCN4 leucine zipper mutants led Harbury et al. (33, 34) and others (38) to certain conclusions about the differences in packing in four-stranded coiled-coils compared with the trimeric and dimeric coiled-coils.

Four-helix Bundles

The four-helix bundle motif, where four α-helices are packed against each other, is frequently found in natural proteins. Such motifs may occur as assemblies of separate polypeptide chains, e.g. repressor of primer, Rop (39), Lac repressor, LacR (40), tumor suppressor, p53 (41), and bacterial luciferase (42); as isolated folds of a single polypeptide chain, e.g. apolipoprotein E3 (43) (Fig. 1A), granulocyte-macrophage colony-stimulating factor (44) (Fig. 1B), human growth hormone (45), and human interleukin-4 (46); or as domains in larger proteins, e.g. T4 lysozyme (47) and 3-isopropylmalate dehydrogenase (48) (Fig. 1C).

Bundles can be grouped into two main classes: where all four helix-helix interaction angles are essentially parallel (Fig. 1A); and where there is a mixture of parallel and perpendicular helix-helix interactions (Fig. 1, B and C). Within these bundle classes, there is potential for considerable interhelical interactions between all four helices. It should be noted that, although 50° is the most commonly observed crossing angle in globular proteins, the 20° angle is more common in fibrous proteins. While a number of different packing angles have been observed, the 50° angle is the most common. A general rule is that if the dimerization angle is larger than 50°, it is likely to be a parallel bundle.
Fig. 1. a–c, 4-helix bundles in apolipoprotein E3 (a), granulocyte-macrophage colony-stimulating factor (b), and 3-isopropylmalate dehydrogenase (c). Two helices are colored yellow and two helices green in b and c. d, dimeric form of 4-helix bundle dimerization domain of aspartate receptor. In one monomer, also shown separately in d, three helices are colored green and one red. In the second monomer, three helices are colored yellow and one blue. The N-terminal helices of each subunit (colored red and blue, respectively) are responsible for dimerization via a two-stranded coiled-coil, shown separately in e, f. Three-stranded coiled-coil domain of influenza hemagglutinin. h and i, DNA binding HTH motif (in yellow) separately (h) and as a homeodomain with a third helix (shown in green) (i). j–l, DNA binding HLH motif as a monomer (l), colored green, as a dimer (h, second monomer colored yellow), and as a bHLH-ZIP protein (l), which contains the HLH dimer shown in k in conjunction with a two-stranded coiled-coil motif (one strand colored red and the other blue). m and n, classical zinc-finger DNA binding motif and β̄βα family variant, respectively. o, GATA-1 transcriptional factor Zn2+ finger. p, first zinc-finger motif in estrogen receptor DNA binding domain, q, "RING finger," which contains two Zn2+ ions in one motif/domain. r, GAL4 zinc-finger motif contains two Zn2+ ions. s, GAL4 dimer consists of two GAL4 motifs dimerized by a two-stranded coiled-coil motif (one monomer in green and the other in yellow); t, single Ca2+ binding HLH or EF-hand motif. u, site III of troponin C, a single Ca2+ binding HLH, which forms a two-site homodimer (one HLH colored yellow and the other green). v, calmodulin Dα, which contains two HLH motifs or EF-hands forming a stable domain (one yellow and the other green). w, N-domain of troponin C also contains two HLH Ca2+ binding sites colored blue and red. x, Ca2+ bound structure of sarcoplasmic calcium-binding protein. The N-domain contains two HLH Ca2+ binding motifs (one red and the other blue). The C-terminal domain contains two HLH motifs (one yellow and the other green).

frequent in four-helix bundles, since it allows pairs of helices to remain in contact over a greater distance (49).

In terms of topology, an up-down-up-down arrangement is the simplest and most common for a four-helix bundle protein (Fig. 1a), with left- and right-turning bundles occurring with approximately equal frequency (49, 50). The cytokine family of proteins is of particular interest as the bundles found in these molecules contain helices arranged in an up-down-down-down topology, which does not exist in any other known protein structures (51).

The simplicity of the four-helix bundle motif has made it ideal as a template for several attempts at de novo protein design (52–54). An interesting variation of this de novo design concept is also offered by the template-assembled synthetic protein approach described by Mutter’s group (55). For more reviews of four-helix bundles, see Refs 49, 56, and 57.

In some cases, protein oligomerization can occur through assembly of single-chain helical bundles into large α-helical assemblies. The crystal structure of the tetrameric enzyme fumarase C from E. coli shows the association of four five-helix bundles, so that the central core of the tetramer consists of 20 α-helices (58). In a similar manner, the aspartate receptor of S. typhimurium and E. coli are dimers, each monomer consisting of an antiparallel single polypeptide chain four-helix bundle (shown in Fig. 1d), where the N-terminal helices of each monomer form the majority of intersubunit contacts, packing together in a parallel coiled-coil dimer interaction (Fig. 1, e and f) with typical 20° crossing angle (59, 60). This protein is a good example of how motifs such as the coiled-coil and four-helix bundle may be integrated to form complex structures involved in protein assembly.

DNA Binding Motifs (bZIP, HTH, bHLH, bHLH-ZIP)

Much recent work has centered on characterizing proteins that bind DNA specifically and control gene expression (61, 62).

The bZIP domain dimerizes through a 30–40-residue coiled-coil. Immediately N-terminal to the coiled-coil is a region rich in basic residues, which is responsible for base-specific DNA binding (61). The crystal structures determined for the GCN4 and Fos/Jun factors bound to their DNA recognition sequences (6, 7, 12) have shown that the coiled-coil dimerization region and the basic sequence form a continuous α-helix that diverges toward its N terminus and passes through the major groove of the DNA binding site. As predicted earlier (63), the bZIP acts like a set of forceps that contacts the dyad symmetric DNA binding site in a “scissors-grip” fashion.

The most common DNA binding motif and the first such motif discovered is the HTH (64). Originally identified in bacterial proteins, it has since been found in hundreds of prokaryotic and eukaryotic DNA-binding proteins (64). The HTH is generally composed of a 20-residue sequence containing two α-helices connected by a turn of usually about 4 residues (Fig. 1h). The helices are about 120° with respect to one another, and the second helix is always the “recognition helix,” which lies in the major groove and participates in base-specific DNA contacts (61, 64). The protein domains that contain the HTH motif can generally be classified into six types according to the identity and positioning of the other structural elements (α-helices, β-strands, or hairpins) that pack against the HTH and close off the hydrophobic core (64). In all of these domains, the HTH motifs show remarkable similarity both in terms of their structure and their mode of binding to DNA through the recognition helix. Recent NMR structure determinations of the ets domain of human FlI-1 (65), LexA repressor DNA binding domain (66), and γ6 resolvase DNA binding domain (67) all agree with the earlier structural results on the HTH motif.

Of particular interest in eukaryotes is the homeodomain, which contains three α-helices including the HTH motif (Fig. 1i) and is generally about 60 amino acids long. Determination of the Antennapedia homeodomain-DNA complex structure by NMR confirmed that the third helix of the homeodomain (helix 2 of the HTH motif) is the DNA recognition helix (68).

One of the most common motifs involved in dimerization and DNA binding of eukaryotic transcription factors is the bHLH motif (15). The motif consists of two amphipathic α-helices joined by an extended loop of between 5 and 24 amino acids (69) (Fig. 1j). The sequence of the bHLH motif is highly conserved among the mem-
Zinc-finger Motifs

The classical zinc-finger motif (80–82) represents a highly conserved class of eukaryotic DNA-binding proteins involved in the regulation of gene expression. This motif of about 30 residues forms an independent monomodule with a single zinc ion tetrahedrally coordinated by 2 cysteine and 2 histidine residues, which give the motif its Cys2His2 nomenclature. This stable monomodule consists of two antiparallel strands of β-sheet connected by a turn, which is stabilized in a helical conformation by an α-helix. The first motif contains 2 histidine ligands, thus forming a ββα fold (Fig. 1a). The most common role of zinc fingers is binding DNA. Of the 10 zinc-finger topologies (81, 82), only five are mentioned here.

The number of repeats of the classical zinc-finger within different proteins ranges from 2 to 37 (81). The three-dimensional structures have been solved by x-ray crystallography for three proteins (containing 2–5 zinc-fingers) complexed with their target DNA binding sites (83, 84). The structures of single or double zinc-finger domains in solution have been determined by two-dimensional NMR spectroscopy in the absence of DNA and are remarkably similar to the crystal structures (85). With an understanding of classical zinc-finger DNA recognition, it was possible to design de novo a zinc-finger protein to recognize an oncogenic sequence (86). Recently, Bianchi et al. (87) used the Cys2His2 consensus zinc-finger motif as a template to display a 5-position library on the surface of the α-helix for rational design of peptidomimetics.

Other variants of the classical zinc-finger have been identified structurally, in which an additional N-terminal β-strand is involved (84) giving rise to a βββα motif (Fig. 1n). GATA-1 is one of a small family of transcriptional factors with a motif that binds a single zinc ion coordinated by 4 cysteine residues; the structure of a 66-residue fragment complexed with DNA was determined by NMR (88) (Fig. 1o). Solution structures of the DNA binding domains of the estrogen and glucocorticoid receptors (89, 90) revealed two zinc binding motifs folded together to form a single structural domain (Fig. 1p). The α-helix of each motif is amphipathic and oriented perpendicularly to the helix of the other motif.

A new structural class of zinc-fingers, referred to as a “RING finger” motif, has been determined by NMR spectroscopy (91) and contains an amphipathic α-helix lying along one surface of a triple-stranded β-sheet (Fig. 1q). Two zinc ions are each coordinated by 4 ligands.

The DNA binding domain of GALA4 represents another motif found in a large group of transcriptional factors (Fig. 1r). The structure of this motif in complex with DNA has been solved by NMR spectroscopy (92, 93) and by x-ray crystallography (17, 18). Six cysteine residues coordinate two zinc ions. The helices are held in a rigid conformation with respect to each other by the sharing of cysteine ligands between the two zinc ions. To build a functional protein entity for binding DNA, GALA4 exists as a dimer where each DNA binding domain is held together by a parallel coiled-coil (Fig. 1s).

Helix-Loop-Helix Ca2+ Binding Motifs

There is a group of some 40 Ca2+-binding proteins, which includes such proteins as troponin C, calmodulin, parvalbumin, calbindin, sarcoplasmic calcium-binding protein (SCBP), calcylin, S100b, recoverin, and α-spectrin (for reviews, see Refs. 94–97). These proteins contain a common structural motif known as the EF-hand or HLH Ca2+ binding motif. This motif generally consists of a 12-residue Ca2+ binding loop flanked by two α-helices (Fig. 1t). The basic structural-functional unit is comprised of a pair of calcium binding sites or EF-hands rather than a single HLH motif (Fig. 1u and v). The pairing of the HLH structures results in a globular domain stabilized by hydrophobic interactions at the interface of the two HLH motifs.

High resolution structures have been determined for a large number of these proteins in either Ca2+-free or Ca2+-bound states by x-ray crystallography (e.g. calbindin D9K (98), calmodulin (99), troponin C (100, 101), SCBP (102), and recoverin (103)) or by NMR spectroscopy (e.g. calbindin D9K (104, 105), homo- and heterodimers of Ca2+ binding sites of the C-domain of troponin C (106–108), N-domain of troponin C (109), C-domain of calmodulin and intact calmodulin (110–112), spectrin (113), S100b (114), calcylin (115), and recoverin (116)). The most in-depth characterization of the structural changes that occur upon binding of metal ions to a 2-site domain has been carried out with the 75-residue calbindin D9K (Fig. 1u).

To understand the Ca2+ affinity and specificity of HLH structures, Hodges and co-workers (117) were the first to take a minimalistic approach by studying a synthetic 34-residue single Ca2+ binding site (site III of troponin C). This peptide formed a symmetric homodimer in a head-to-tail arrangement in the presence of Ca2+ (Fig. 1u) (106). Similarly, a 39-residue proteolytic fragment containing Ca2+ binding site IV of troponin C was shown to form a dimer (108). These and other (107, 118) studies have clearly indicated that dimerization of single HLH structures controls Ca2+ affinity and that even the homodimers bind two calcium ions with positive cooperativity. Clearly, the detailed hydrophobic interactions in the interface between calcium binding sites stabilize the domain and control Ca2+ affinity (119).

The two-site HLH domain is the minimum stable folding unit, and in molecules like troponin C and calmodulin, two of these two-site domains form an extended dumbbell-shaped structure (Fig. 1v). In contrast, molecules like SCBP (102) and recoverin (103), which also contain 4 HLH motifs, bring the N-terminal and C-terminal halves of the protein together to form a highly compact and globular structure (Fig. 1x).

The structure of apo-calcyclin, an S100 calcium-binding protein, reveals a homodimERIC structure where each polypeptide chain of approximately 90 residues contains two HLH motifs (115, 120). Shaw and co-workers (114) prepared a synthetic peptide, residues 1–46 of human brain S100b, analogous to the first HLH of calcyclin. This peptide assembles into a tetramer in the presence of Ca2+. Finally, the HLH motif, depending on its sequence, can assemble into two- or four-site domains with varying Ca2+-dependent regulatory or Ca2+-buffering functions.

In conclusion, understanding the details of folding and stability of the assembly motifs discussed in this review should enable the de novo design of novel proteins referred to as “hybridins,” consisting of multiple and different motifs joined together in the same polypeptide chain and which interact in a cooperative manner (121).

REFERENCES

1. Crick, F. H. C. (1953) Acta Crystallogr. 6, 689–697
2. Adamson, J. G., Zhou, N. E., and Hodges, R. S. (1993) Curr. Opin. Biotechnol. 4, 426–437
3. Hodges, R. S. (1992) Curr. Biol. 2, 122–124
4. Hodges, R. S. (1996) Biochem. Cell Biol. 74, 133–154
5. O’Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) Science 254, 339–344
6. Glover, J. N., and Harrison, S. C. (1995) Nature 373, 257–261
7. Ellenberger, T. E., Brandl, C. J., Struhl, K., and Harrison, S. C. (1992) Cell 71, 1223–1237
8. Harborth, J., Weber, K., and Osborn, M. (1995) EMBO J. 14, 2447–2460
9. Cohen, C., and Parry, D. A. D. (1994) Science 263, 488–489
10. Jones, N. (1990) Cell 61, 9–17
11. Baxevanis, A. D., and Vinson, C. R. (1993) Curr. Opin. Genet. Dev. 3, 278–285
12. Konig, P., and Richmond, T. J. (1993) J. Mol. Biol. 233, 139–154
13. Atkinson, R. A., Saudek, V., Huggins, J. P., and Pelton, J. T. (1991) Biochemistry 30, 9387–9395
14. Sundek, J., Bohm, A., Lambricht, D. G., Hamm, H. E., and Sigler, P. B. (1996) Nature 379, 369–374
