Extended knobs-into-holes packing in classical and complex coiled-coil assemblies

John Walshaw\(^1\) and Derek N. Woolfson\(^*\)

Centre for Biomolecular Design and Drug Development, John Maynard-Smith Building, School of Life Sciences, University of Sussex, Falmer BN1 9QG, UK

Received 1 October 2003, and in revised form 16 October 2003

Abstract

This year marks the 50th anniversary of Crick’s seminal paper on the packing of \(z\)-helices into coiled-coil structures. The central tenet of Crick’s work is the interdigitation of side chains, which directs the helix–helix interactions; so called knobs-into-holes packing. Subsequent determinations of coiled-coil-protein sequences and structures confirmed the key features of Crick’s model and established it as a fundamental concept in structural biology. Recently, we developed a program, SOCKET, to recognise knobs-into-holes packing in protein structures, which we applied to the Protein Data Bank to compile a database of coiled-coil structures. In addition to classic structures, the database reveals 4-helix bundles and larger helical assemblies. Here, we describe how the more-complex structures can be understood by extending Crick’s principles for classic coiled coils. In the simplest case, each helix of a 2-stranded structure contributes a single seam of (core) knobs-into-holes to the helical interface. 3-, 4-, and 5-stranded structures, however, are best considered as rings of helices with cycles of knobs-into-holes. These higher-order oligomers make additional (peripheral) knobs-into-holes that broaden the helical contacts. Combinations of core and peripheral knobs may be assigned to different sequence repeats offset within the same helix. Such multiple repeats lead to *multi-faceted helices*, which explain structures above dimers. For instance, coiled-coil oligomer state correlates with the offset of the different repeats along a sequence. In addition, certain multi-helix assemblies can be considered as *conjoined* coiled coils in which multi-faceted helices participate in more than one coiled-coil motif.

© 2003 Elsevier Inc. All rights reserved.

Keywords: Fibrous proteins; Francis Crick; Helix–helix packing; Helical bundles; Heptad repeats; Protein–protein interactions

1. Introduction

A major question in protein folding is how does sequence dictate three-dimensional structure? Relationships between protein primary and secondary structures are becoming well established. This leaves the bigger hurdle of understanding how secondary structures pack to form tertiary and quaternary structures. In this respect, coiled-coil motifs present an intriguing conundrum. Coiled coils are \(z\)-helical assemblies found in proteins and at protein–protein interfaces. Although more-complex sequence patterns do occur (Hicks et al., 2002; Lupas, 1996), the accepted hallmark of coiled-coil sequences is the heptad repeat (Crick, 1953; Lupas, 1996), which is a contiguous run of a 7-residue consensus pattern of hydrophobic (H) and polar residues (P), HPPHPPP. This repeat encodes amphipathic \(z\)-helices that assemble into bundles via their hydrophobic faces. The puzzle is that similar patterns lead to a variety of different three-dimensional structures: classic coiled coils have between two and five helices arranged in parallel, anti-parallel or mixed topologies (Lupas, 1996); heptad repeats also appear in \(z\)-helical domains that are more complex than classic coiled coils (Walshaw and Woolfson, 2001a,b).

By convention, the residues of the heptad are labelled \(abcdefg\). Hydrophobic residues tend to occupy the \(a\) and \(d\) sites so that in a repeating sequence they are...
alternately spaced three and four positions apart. Thus, when configured into an α-helix, which has 3.6 residues per turn, the a and d residues are brought together to set up a hydrophobic seam. However, because seven residues fall short of two complete α-helical turns the seam drifts around the helix surface with the opposite hand to the helix itself; i.e., the seam is left-handed. Therefore, in order for the hydrophobic seams of two or more such helices to marry, the helices must wrap, or coil around each other in a left-handed manner (Fig. 1A). The helices are said to supercoil. All slices through this supercoil comprising one heptad from each helix are equivalent; the two helices have the same orientation with respect to each other at any point along the supercoil. Therefore in 'supercoil space,' a coiled-coil helix has exactly seven residues every two turns (Fig. 1B).

In terms of theories for helix packing, coiled-coil interactions fit with the earliest and most extreme model. This is the knobs-into-holes system proposed by Crick 50 years ago (Crick, 1953). Other theories have since been developed to explain helix packing in general. These include models of ridges-into-grooves (Chothia et al., 1981) and interlacing-of-Cα-positions (Walther et al., 1996). The extent to which these models are realised in globular helical bundles, however, has been questioned (Bowie, 1997). Furthermore, although the interlacing of Cα positions may be a general feature of helix packing (Walther et al., 1996), the degrees to which the side chains of neighbouring helices interdigitate varies (Chothia et al., 1981; Efimov, 1999; Walshaw and Woolfson, 2001b). By contrast, it is commonly accepted that the knobs-into-holes description with its tight side-chain interdigitation, is appropriate for explaining the packing of helices in coiled-coil motifs (Bowie, 1997; O’Shea et al., 1991).

Specifically in coiled coils, hydrophobic side chains at a and d on one helix act as knobs and dock into holes formed by diamonds of four residues on a partnering helix. For parallel helices, a knobs dock into holes formed by d+1gad residues; d knobs interact with adea+1 holes. (Subscripts ±1 refer to positions in heptads following and preceding the heptad in which the knob residues fall.) The resulting helix packing is extremely tight-knit (Fig. 2).

These sequence and structural features first described by Crick (1953) were confirmed with the first sequences and X-ray crystal structures of coiled-coil proteins (Stone et al., 1975; Wilson et al., 1981), and more recently with high-resolution structures of coiled coils such as the leucine zipper (Fig. 1) (O’Shea et al., 1991) and its variants (Harbury et al., 1993, 1994). In testament to Crick’s forward-thinking and structural insight, the crystal structures highlight impeccably the helical supercoiling and inter-helix knobs-into-holes packing of side chains that he predicted. The leucine-zipper motif has since become the archetypal example of the coiled coil, and the favoured model system for study. Of particular relevance is the work of Harbury and colleagues, which shows that the geometries of knobs packing into holes differ in dimers, trimers, and tetramers (Harbury et al., 1993, 1994). Moreover, these different geometries lead to different amino-acid preferences for the various knob positions (Harbury et al., 1993; Woolfson and Alber, 1995). Thus, albeit in a relatively simple system, oligomer state can be selected with some certainty by combining particular residues at the a and d sites of the heptad repeats (Harbury et al., 1993, 1994; Nautiyal and Alber, 1999; Nautiyal et al., 1995; Pandya et al., 2000).

Considerable effort has been placed on characterising known coiled-coil sequences and devising algorithms that recognise coiled-coil motifs in primary structure.
1. **Highly symmetric, classic coiled coils.** These are fibrous domains usually at least four heptads long, and almost exclusively based on canonical heptad repeats. However, some shorter domains are present, as are motifs containing non-canonical, 11-residue repeats (Hicks et al., 1997, 2002; Lupas, 1996).

2. **Irregular and short pairs of adjacent helices.** These are usually in the context of a larger globular α-helical domain. Despite their apparent lack of symmetry, the packing in these helix-pairs is hard to distinguish from long, fibrous, 2-helix coiled coils (Walshaw and Woolfson, 2001b). This category also includes examples from several membrane-spanning domains.

3. **4-Helix bundle domains.** These are distinguished from 4-stranded coiled coils as they have one or more pairs of helices with coiled-coil interactions rather than complete cyclic knobs-into-holes packing characteristic of classic coiled-coil tetramers (see below); the remaining helices do interact with each other, but in a non-coiled-coil manner.

4. **Largely symmetric, multi-helix assemblies.** These include a variety of arrangements comprising three or more helices and display tight, regular knobs-into-holes interactions, but, at a glance, are not classic coiled coils.

Domains of type 1 are well known in the literature (Lupas, 1996), and we have discussed them along with those in category 2 at length elsewhere (Walshaw and Woolfson, 2001b). Here, we explain that assemblies in groups 3 and 4 can be explained through straightforward extensions to Crick's structural principles for classic coiled-coil dimers and trimers, hence the term “extended knobs-into-holes packing” in our title. Characterisation of the regular packing principles of the domains of type 4 may enable recognition from amino-acid sequence of novel α-helical domains.

2. **Discussion**

The following discussion is illustrated by reference to the structures and SOCKET outputs (Appendix) for the dimeric leucine zipper peptide, GCN4-p1 (PDB code 2zt; O'Shea et al., 1991) its trimeric (pII, 1gcm; Harbury et al., 1994) and tetrameric (pLI, 1gel; Harbury et al., 1993) variants, and a 5-stranded coiled coil (Malashkevich et al., 1996). These are parallel structures. However, the general concepts may be extended to anti-parallel coiled coils though the precise details will necessarily differ. This is because: (1) the residues involved in the knobs-into-holes and (2) the packing angles that the knobs make with the holes are different between parallel and anti-parallel structures (Walshaw and Woolfson, 2001b).

2.1. **The concept of a half-seam**

As expected, the SOCKET output for the 2-helix coiled coil, GCN4-p1, reveals that the *a* and *d* residues of the *abcdefg* heptad repeat forms a seam of knobs on each helix (Appendix). The two seams packs together to form the hydrophobic interface. This is depicted in Fig. 2 where the seams are emphasised by discs. For reasons that will become clear, it is useful to introduce the concept of the “half-seam.” By this we mean a subset of knob residues, for example, all of the *a* residues from one helix. A half-seam is equivalent to one of the lanes bounded by the dotted lines on the surfaces of the helices of Fig. 2. Again as expected for GCN4-p1, SOCKET shows that the two half-seams from one helix—one of *a* residues the other of *d*—plug into the same neighbouring helix; there is no other choice. In this case, the combined interface of two half-seams is a traditional single seam. This is illustrated as a helical-wheel diagram in Fig. 3A.

The cartoons of Figs. 4A and B decompose the heptads to separate the *a* and *d* knobs and show that they fall in distinct layers; a layer comprises residues at approximately the same level on the coiled-coil axis; for example, *g* & *a* and *d* & *e*. These coincidences occur because the rise per residue along the helix axis is roughly countered by the leaning of the helices with respect to the coiled-coil axis, which is due to the supercoiling. Layers are roughly perpendicular to the coiled-coil axis and form the basis of the knobs-into-
holes interactions; for example, a knobs always fit into holes formed by g & a residues of the same layer in a neighbouring helix, the remaining sides of the hole are provided by d residues from the layers above and below. Furthermore, in 2-stranded coiled coils, like GCN4-p1, knob residues also form one side of the hole into which the other knob from the same layer (i.e., that from the interacting helix) fits. We refer to this as an arrangement of pairwise-complementary knobs (Walshaw and Woolfson, 2001a,b).

Pairwise complementarity only holds for 2-stranded coiled coils. As illustrated below, in the 3- and 4-stranded variants of GCN4 the a and d half-seams plug into different neighbouring helices, which leads to cyclic complementarity of knobs-into-holes (Walshaw and Woolfson, 2001a,b); incidentally, cyclic knobs-into-holes packing in “3-strand ropes” was also anticipated by Crick (1953).

This splitting of the a plus d interface in the 3- and 4-helix structures, pII and pLI, can be seen in the helical wheels of Figs. 3B and C and the SOCKET outputs (Appendix). Again, however, it is much clearer in the decomposed layer diagrams of Figs. 4C–F. For example in the trimer, the a side chains of helix X dock with helix Z, whereas those at d interface with helix Y (Figs. 4C and D, Appendix). Furthermore, whilst the a knobs cycle within the gla layers in one direction (Fig. 4C), the d knobs cycle in the opposite sense (Fig. 4D). This same relative orientation of the a and the d knobs is maintained in the tetramer (Figs. 4E and F; Appendix).

2.2. *The concept of peripheral knobs and holes*

The SOCKET analyses of GCN4-p1, pII, and pLI (Appendix) reveals another feature of coiled-coil assembly that alters with the step up from 2-stranded to higher-order structures. For the latter, the traditional 3,4-spacing of knobs is supplemented; in other words, SOCKET identifies knob residues outside the canonical a and d positions.

For the trimer, pII, there are two additional knobs at e positions. However, in the tetramer, pLI, complete, or partially complete additional half-seams are identified at the e and g sites for all four helices. These interface with ‘holes’ formed partly by cd and ab residues, respectively, on neighbouring helices (see Fig. 2). Furthermore, the e half-seam plugs into the same helix as the a half-seam; and the g half-seam plugs into the same helix as the d half-seam. In other words, the a plus e half-seam makes one collective interface, and d plus g another. This arrangement is summarised in Fig. 5D, and is evident in the SOCKET output (Appendix).

We refer to the knobs formed outside the ald core as peripheral knobs, and the associated holes as peripheral holes. As seen in the SOCKET output for pLI, peripheral interactions are less geometrically ideal than core
interactions, and are not always found by SOCKET; for example, contrast the numbers of peripheral knobs located for each of the helices in the pLI tetramer (see supplementary material). Nonetheless, we find that peripheral knobs are a general feature of coiled-coils with more than two strands.

2.3. The number of peripheral knobs-into-holes increases with oligomer order

Why are the peripheral interactions observed in the pLI tetramer mostly absent from the pII trimer? The reason is that in lower-order assemblies neighbouring helices interact at much steeper angles than in the

---

Fig. 8. (A,C) Molscript (Kraulis, 1991) ribbon diagrams depicting knob side chains determined by SOCKET and (B,D) schematic cross-sections of conjoined coiled coils; thick lines between helices indicate trimeric, cyclically complementary, knobs-into-holes packing. (A,B) Stat3β (1bg1; Becker et al., 1998): helices α2 and α3 contribute to two conjoined anti-parallel 3-stranded coiled coils; (C,D) HIV gp41 ectodomain (1aik; Chan et al., 1997): each N36 helix is simultaneously part of three different 3-stranded coiled coils (one parallel and two anti-parallel).

---

Fig. 9. (A) Molscript (Kraulis, 1991) ribbon diagram of the aspartate receptor (1vlt; Yeh et al., 1996) depicting knob side chains determined by SOCKET. (B) Schematic cross-section of (A); thick lines between helices indicate both the pairwise and cyclically complementary knobs-into-holes packing. The A helices form a 2-stranded coiled-coil with each other, and a 3-stranded coiled coil with helices B and C of the same chain.
higher-order structures, Fig. 3. As a result, as oligomer state increases side chains at e and g make more-intimate contacts with the neighbouring helices.

For example, in 2-stranded structures the rotational offset between the helices is as steep as possible, 180°; consistent with this, peripheral knobs are not seen in the GCN4-p1 dimer (Appendix), and we rarely observe them in other 2-stranded structures of the whole database. Even with the rotational offset reduced to 120° in pII and other 3-helix assemblies, peripheral knobs are rarely detected because the e and g knobs are usually still not close enough to dock into the potential cd and ab holes. There are exceptions, however, and SOCKET detects some peripheral knobs-into-holes in 3-stranded structures; these necessarily involve long side chains in conformations that allowed interactions with the respective holes, as seen in helices Y and Z of the pII trimer (Appendix). The increased numbers of peripheral knobs in pLI and other tetramers reflects the reduction of the offset to 90°, which permits peripheral interactions almost regardless of the nature of the side chain. This trend continues in rat COMP, a homopentameric coiled-coil structure (Malashkevich et al., 1996) for which SOCKET detects numerous peripheral knobs-into-holes. Presumably this results from the even shallower rotational offset of 72°. However, the COMP structure also illustrates that the docking of e and g residues is achieved partially at the expense of the core packing of a and d residues. This can be understood as follows:

Imagine the helices of a dimeric coiled coil; that is, two helices rotated by 180° with respect to each other and with their ald core residues directly facing one another (Fig. 3A). Introducing more helices broadens the helix–helix interfaces, lowers the rotational offset between helices, and leads to the recruitment of peripheral knobs-into-holes interactions. However, to maintain coiled-coil symmetry, it also requires each helix to rotate around its own axis (Figs. 3B and C). Thus for higher-order assemblies, the ald seams no longer face each other, but are directed more towards the centre of the assembly (Fig. 3). In trimers and tetramers, and with the correct choice of side chains at a and d, this indirect orientation of interfaces can result in better packing (Harbury et al., 1993). However, in pentamers a consequence of the smaller rotational offset between neighbouring helices is that the a and d side chains fit rather shallowly into their ga and de holes and packing is compromised.

Incidentally, on this basis true 6-stranded coiled coils are improbable, because the rotational offset of 60° would lead to an unfavourably shallow angle for core residues to fit into their holes. We note that higher-order, symmetrical arrangements do exist, but necessarily are not based on coiled-coil interactions. For instance, modification of a designed 2-stranded coiled coil, by replacement of charged groups with alanine to abolish interactions
between the e and g positions, results in the formation of a regular octameric bundle without knobs-into-holes packing or coiled-coil specificity (Meier et al., 2002).

2.4. Offset double heptad repeats

How do peripheral knobs tally with the accepted signature of coiled-coil sequences, namely the heptad repeat? The SOCKET output for the pLI tetramer sheds light on this (Appendix). Consider the completed, consensus output of knobs located by SOCKET for one of the helices of pLI, for example W, which directs knobs to helices X and Z (Fig. 5D):

This can be decomposed into the interface with helix Z:

This can be decomposed into the interface with helix X:

In both decompositions the spacing of knob residues is 3,4; i.e., as in a canonical heptad repeat for a 2-helix coiled coil. Thus, two heptad repeats are effectively superimposed on the same sequence. In this case, the two repeats are offset by one residue.

Comparing line (1) with lines (2) and (3) shows that we reassigned the heptad registers for the decomposed interfaces to place knobs at the traditional a and d sites. In the first decomposition a residues become d, and e become a; whilst in the second, d residues become a, and g become d. In other words, the a+e interface of the tetramer (line (1)) appears similar to a d+a interface of a normal dimeric coiled coil (line (2)); and the d+g interface (line (1)) similar to an a+d interface (line (3)). This can be visualised by comparing Figs. 3A and 5D, which show how the tetramer d residues are structurally similar to dimer a residues, and that the tetramer g residues play the role of the dimer d residues. This reassignment of the heptad registers for the decomposed interfaces is useful because it is another way of expressing the Harbury rules for core-packing geometry in dimeric and tetrameric coiled coils (Harbury et al., 1993); that is, that the core packing geometries at a and d in dimers are effectively swapped in tetramers.

Fig. 4. Helical-wheel schematics of a-layers (A, C, and E) and d-layers (B, D, and F) in 2-stranded (A and B), 3-stranded (C and D), and 4-stranded coiled coils (E and F).
Thus, dimeric and tetrameric coiled coils are related. However, the interfaces are not completely identical because in a dimer, the two helices have second-order rotational symmetry and the hydrophobic seams face each other. In the tetramer, which has fourth-order rotational symmetry, each helix is tilted slightly about its axis, which effectively points the $a$ and $d$ residues more towards the centre of the structure. This moderates the knobs-into-holes packing as described above; n.b. above, we are comparing a pair of helices in a dimer, with a pair in a tetramer whose heptad register has been reassigned. If instead we had compared the dimer and the real heptad of the tetramer, then there is a much larger tilt, $\sim 45^\circ$, as noted by Harbury et al. (Harbury et al., 1993).

Summarising this section, SOCKET identifies $e$ and $g$ residues as knobs in structures such as pLI because they are almost structurally identical to the traditional $a$ and $d$ core residues of dimers such as GCN4-p1. Combinations of peripheral and core knobs can be reassigned to independent heptad repeats superimposed on the same stretch of sequence.

Incidentally, SOCKET analysis of the pentameric coiled coil, COMP (Malashkevich et al., 1996), similarly reveals two sets of knob residues for each helix (Appendix). These can also be decomposed as superimposed heptad repeats offset by one residue in the sequence. In geometrical terms, this particular offset is compatible with both helical tetrators and pentamers because of

---

Fig. 5. Helical-wheel schematics for the variations of knobs-into-holes observed in 4-helix bundles and 4-stranded coiled coils. Key: Circles with dotted centres are $C^a$ atoms. White circles show $C^a$ atoms that form the sides of a hole, but are not knobs. Black circles represent $C^a$ atoms for non-complementary knobs. Grey circles show $C^a$ atoms of complementary knobs. Grey and black knobs are in some cases part of a hole. (A) Layer from a 4-helix bundle with no true knobs-into-holes packing; (B) layer from a 4-helix bundle in which helix $Z$ is part of two 2-stranded knobs-into-holes interactions with helices $W$ and $Y$; (C) $a$-layer from a 4-stranded coiled coil with peripheral as well as core knobs-into-holes; (D) as (C), but showing both $a$- and $d$-layers.
the \(103^\circ\) angular offset between the \(a/e\) and \(d/g\) faces, Fig. 6. However, at present we do not understand the sequence features of pLI and COMP that lead to the different oligomers in these cases; although an explanation has been provided for the selection of the pentameric form of COMP (Malashkevich et al., 1996).

### 2.5. 2-Helix coiled coils within 4-helix bundles

The heptad reassignment described above for 4-stranded coiled coils is also useful in distinguishing these structures from 4-helix bundles that have only partial coiled-coil character; for instance, where only one or two pairs of helices are like 2-stranded coiled coils, four-fold symmetry is lost and the helices do not tilt towards the centre of the assembly.

For example the ferritin structures have two helix-turn-helix motifs that pack anti-parallel alongside each other. SOCKET analysis of bullfrog L-ferritin (1rcc; Trikha et al., 1995) and bacterioferritin (1bcf; Frolow et al., 1994) indicates only pairwise-complementary knobs-into-holes interactions within the helix-turn-helix motifs, and no cyclic knobs-into-holes interactions (Fig. 7, Appendix). Therefore, this fold is not a 4-stranded coiled coil.

Closer analysis of the ferritins reveals that each of the 2-stranded coiled coils has tight core-packing, with the hydrophobic interfaces directly facing each other as in a classic 2-helix coiled coil. On the other hand, no knobs-into-holes packing at all occurs between any of the helices of the neighbouring coiled coils: in 45 domains that we identified with this fold, the mean distance between the helices within the N-terminal and C-terminal coiled coils were 7.9 and 8.4 Å, respectively; whereas, the mean distances between the adjacent helix pairs with no knobs-into-holes packing increased to 9.7 and 11.3 Å. The two coiled coils therefore pack against each other in a non-coiled-coil fashion. Interestingly, we found that the packing within the coiled coils was tighter and more regular in the bacterioferritins than the vertebrate ferritins: all the bacterioferritins had tight knobs-into-holes packing along almost the full length of each pair of helices; whereas, the vertebrate ferritins had smaller and more variable numbers of complementary layers (http://www.biols.susx.ac.uk/coiledcoils/ccc/cat/mulal/famE_ferritin_F.html).

We found other 4-helix bundles to be more complicated and less regular. For instance, many cytokines exhibited pairwise knobs-into-holes packing interactions, but no cyclic interactions. Although these did not resolve as coiled coils, some individual helices had characteristics of a helix in a true 4-helix coiled coil; i.e., as illustrated in Fig. 5B, the helices had two heptads offset by one amino-acid position. SOCKET identifies an example of this in human growth hormone (1a22; Clackson et al., 1998), Appendix.

A detailed analysis of knobs-into-holes packing in 4-helix bundles is beyond the scope of this discussion. However, it is noteworthy that the sequences of regions identified by SOCKET as having complementary knobs-into-holes are in some cases also identified by sequence-based coiled-coil prediction programs. For example, many of the helices in the growth hormones and ferritins (but, interestingly, not the bacterioferritins) score highly using the 14- or 21-residue window of COILS (Lupas et al., 1991). High scores with the default, 28-residue window were also noted for a few structures, e.g., 1huw (Ultsch et al., 1994). The partial coiled-coil nature of such sequences, demonstrated here at the structural
level, illustrates that these high scores should not necessarily be dismissed as false-positives.

2.6. Other offsets of double heptad repeats

So far we have considered helices with two interfaces arising from two heptad repeats offset by one residue. Each interface packs with a different neighbouring helix, and comprises one core residue (a or d) and usually one peripheral knob (e or g). For 7-residue repeats, there are two other possible sequence offsets of double heptads, namely 2- and 3-residue offsets; n.b., 1-, 2-, and 3-residue offsets are equivalent to offsets of 6, 5, and 4 positions, respectively, Fig. 6.

The 2-residue offset differs from the other arrangements because it results in two distinct seams on opposite sides of a helix. Such helices may, therefore, participate in two fully independent interactions; indeed, oligomerisation of similar helices leads to α-sheet or α-cylinder structures. Along with others, we have described this offset and the assemblies that it leads to elsewhere (Calladine et al., 2001; Koronakis et al., 2000; North et al., 2001; Walshaw and Woolfson, 2001a).

The 3-residue offset is manifest in trimeric coiled coils. For instance, SOCKET analysis of the pII trimer shows that each helix also display two interfaces, which again dock with different neighbouring helices (Appendix). However, the interfaces comprises only a and d knobs and peripheral e or g knobs are rarely identified. This is shown in the decomposition for helix X of pII, which interfaces with helices Y and Z:

abcdefgabcdedefgabcdedef

---Y-----Z-----Z---Y-----Z---Y---

Interface with Z:

abcdefgabcdedefgabcdedef

-------Z-------Z-------Z-------

Interface with Y (register reassigned):

efgabcdefgabcdedefgabc

---Y-------Y-------Y-------Y---

Thus, each interface comprises a single half-seam, and these two half-seams are offset by three positions. Figs. 4C and D illustrate this graphically.

Consider what would happen if these two half-seams were completed to make two full seams: simply add Z's at the d positions of the Z interface, and Y's at d in the Y interface. Superimposing these hypothetical sequences back onto the united (non-decomposed) register would augment the pattern with new Y knobs at g. However, in addition, the new Z knobs would coincide with the original Y knobs at d. This would require these side chains to fit simultaneously into holes on two different helices, which is highly unlikely. Thus, such full-seam
interfaces are not expected in trimers nor did we observe them in the SOCKET analysis of the entire PDB.

For 3-, 4-, and 5-stranded coiled coils, our purpose in decomposing the united registers into its component parts is to highlight the hierarchy in these structures, and to show how separate helix–helix interfaces relate to and may be understood in terms of 2-helix coiled-coil interactions. However, neither of the component registers is the traditional register, which describes the central core. Indeed, the knob residues of double heptad sequences with 1- and 3-residue offsets can always be united as a single heptad with a single, central core of cyclic layers of either \( a \) knobs or \( d \) knobs, which, of course, is how the heptad registers of such structures are traditionally considered.

2.7. Multiple offset repeats and more-complex coiled-coil assemblies

Thus far we have considered only bi-faceted helices. What happens if there are more than two offset hydrophobic repeats? Our SOCKET analysis of the PDB reveals such structures. All were based on trimeric kernels.

2.7.1. Stat3\( \beta \): a pair of conjoined trimers

For example, the mouse Stat3\( \beta \) homodimer contains helices with three interfaces offset by three heptad positions (Becker et al., 1998). By sight, the Stat3\( \beta \) N-terminal domain appears to be a 4-helix bundle. SOCKET shows that it is not a 4-helix coiled coil. However, the analysis demonstrates that two of the four helices, \( \alpha_2 \) and \( \alpha_3 \), are simultaneously part of two different anti-parallel 3-helix coiled coils (Figs. 8A and B and Appendix). We describe this arrangement as a pair of conjoined coiled-coil trimers that share two helices. As we discuss elsewhere (Walshaw and Woolfson, 2001a), helix sharing has implications for coiled-coil supercoiling, and such arrangements could not extend very far; indeed, there are only two complete layers of knobs in the inclusive conjoined arrangement.

2.7.2. gp41: a conjoined trimer of trimers

The ‘most stripy’ helices are found in certain viral membrane-fusogenic domains with six helices. Chan and colleagues describe an assembly of two peptides corresponding to the N- and C-terminal \( \alpha \)-helices of the ectodomain of the HIV-1 envelope glycoprotein gp41 (1aik; Chan et al., 1997). The 36-residue N-terminal peptide (N36) forms a parallel 3-stranded coiled coil, while three copies of the 34-residue C-peptide (C34) pack anti-parallel into the outer grooves of the N36 trimer, Figs. 8C and D. Although it is possible to assign a heptad register to the C34 sequence, the authors of the structure suggest that the outer triplets—comprising two copies of N36 and one of C34—should not be considered as coiled coils because they certainly lack the symmetry of the inner trimer (Chan et al., 1997). Indeed, the C34 helices are significantly tilted with respect to the inner N36 structure. However, SOCKET identifies four trimeric coiled-coil units within the assembly (Appendix), with each based on true knobs-into-holes interactions with cyclic complementarity: firstly, SOCKET confirms the inner assembly as a parallel trimer of N36; secondly, it highlights five consecutive mixed \( ald \) layers of knobs and holes in each of the outer structures and assigned them as anti-parallel trimers. Note that this arrangement is essentially the same as the Stat3\( \beta \)-N-terminal domain, but with two additional helices (compare Figs. 8B and D).

We find several other viral fusogenic proteins that exhibit this pattern of conjoined coiled coils, notably, the SIV gp41 ectodomain (1qbt; Yang et al., 1999) and the Visna virus membrane fusion protein (1jk; Malashkevich et al., 2001); the term ‘trimer-of-hairpins’ has been suggested as a general name for these motifs (Malashkevich et al., 2001). An exception is the Ebola virus gp2 peptide (2ebo; Malashkevich et al., 1999); in this case, the central three helices resolve as a close-packed coiled coil, but the outer helices only exhibit 2-stranded interactions with the central helices.

The N36 peptides of gp41 each make knobs-into-holes interactions with four other helices and, therefore, have four half-seams, which make it the most-stripy helix that we find in the PDB. However, with many of the extended knobs-into-holes interfaces that we describe, these seams are not exclusively formed by hydrophobic residues.

2.7.3. Aspartate receptor: two side-by-side 4-helix bundles hide a dimer of 3-helix coiled coils

The ligand-binding domains of the aspartate receptors involved in bacterial chemotaxis form dimers. Each chain is a 4-helix bundle. Within these bundles, SOCKET reveals anti-parallel 3-helix coiled coils cemented by knobs-into-holes interactions along stretches of the first three helices. For instance, in the representative structure (1vlt; Yeh et al., 1996) this involves residues 56–70, 90–103, and 131–141. The fourth helix (145–172) is not part of any coiled coil. The dimer interface also involves the first helix, which contributes additional knobs and holes between residues 58 and 72. Thus, the first helix is shared between two coiled-coil systems, Fig. 9. This is clear from the SOCKET output for this helix, which has two entries: one for a parallel 2-stranded coiled coil and one for a 3-stranded anti-parallel coiled coil (Appendix). Accordingly, the sequence of these helices shows two superimposed heptad patterns. Based on the heptad assignment for the parallel dimer, residues at \( a \) and \( d \) make up one complete seam and form the homodimer interface. Offset from this, two half-seams plus some peripheral knobs contribute to the cyclic knobs-into-holes pattern of the 3-stranded coiled
coiled coil; these are made up of residues at $b$, $c$, and $f$ of the heptad assignment for the dimer. Thus, this arrangement spreads the dimer and trimer interfaces as far from each other as possible.

3. Conclusions

The coiled coil is probably the most ubiquitous protein–protein interaction motif; 5–10% of all gene sequences are estimated to encode coiled-coil regions (Newman et al., 2000; Walshaw and Woolfson, unpublished). Traditionally, coiled coils are identified at the sequence level by searching for the coiled-coil hallmark; namely, the 7-residue (heptad) repeat of hydrophobic (H) and polar (P) residues, HPPHPPP. However, we find that the favoured sequence-analysis tools disagree on the location and extent of coiled-coil regions in specific genes and whole genomes (Walshaw and Woolfson, unpublished). This presents problems for predicting, analysing and engineering coiled-coil motifs.

To begin bridging this sequence-structure gap we returned to Crick’s original postulate for structural basis of coiled-coil structure—namely, that supercoiled helix–helix interactions are cemented by knobs-into-holes packing of side chains (Crick, 1953)—and developed SOCKET to recognise coiled coils at the structural level (Walshaw and Woolfson, 2001b). We used SOCKET to build a database of coiled-coil structures (http://www.biols.susx.ac.uk/coiledcoils/).

SOCKET identifies the characteristic structural feature of coiled coils; namely, the knobs-into-holes interdigitation of hydrophobic residues. When applied to the Brookhaven Protein Data Bank, SOCKET locates classic 2-, 3-, 4-, and 5-stranded coiled coils and various new structures. Inspection of the classic structures showed that 2-stranded structures are adequately described by the traditionally accepted model: with the positions of the heptad labelled abcddefg, the predominantly hydrophobic $a$ and $d$ residues of one helix form knobs that combine with diamonds of residues (holes) on the neighbouring helix. In the higher-order, 3-, 4-, and 5-stranded structures, however, the helix–helix interfaces are broadened. This is not in a non-specific, or general sense. Rather, the number of knobs-into-holes interactions increases with positions $e$ and $g$ contributing knobs to augment the traditional heptad pattern. The number of these peripheral knobs increases with oligomer state.

These extended knobs-into-holes interfaces can be rationalised in terms of multiple repeats of knob residues offset along the same sequence. For instance, 4- and 5-stranded structures are based on two heptad repeats offset by a single residue; in trimers the offset is three residues. The only other possible offset is of two residues, which leads to novel $\alpha$-sheet and $\alpha$-cylinder structures (Calladine et al., 2001; Koronakis et al., 2000; North et al., 2001; Walshaw and Woolfson, 2001a). Multiple patterns of knobs offset on a single helix are also possible. These lead to complex coiled-coil assemblies in which one or more helices are effectively shared by different coiled-coil units. Examples include conjoined trimers in Stat3\(\beta\), a dimer of trimers in the aspartate receptor and a trimer of trimers in the HIV-1 envelope protein gp41.

Our observations and discussion raise two questions, however:

First, how do multi-faceted coiled-coil structures tally with the accepted heptad signature of hydrophobic and polar residues, HPPHPPP, of coiled-coil sequences, which gives only one hydrophobic seam? Attempts are being made to answer this question (Calladine et al., 2001; Walshaw and Woolfson, 2001a), but a detailed answer will require in-depth analyses of the sequences and structures for families of multi-faceted coiled-coil structures. In the meantime, we expect that modifications of the classical pattern will be necessary to facilitate multiple, offset seams. Indeed, even amongst the conventional 2-helix coiled coils in our previous analysis (Walshaw and Woolfson, 2001b) we found very frequent exceptions to the HPPHPPP template.

Second, how do shared helices supercoil—a requirement for heptad-based knobs-into-holes packing—simultaneously in two or more coiled coils? From our observations so far (Walshaw and Woolfson, 2001a), it appears that the shared helices do this: (1) by distorting and specifically by straightening, and (2) by making multi-faceted interactions for only short stretches. The second point necessarily follows from the first because distortions cannot be maintained indefinitely. In this sense, such arrangements cannot be considered true coiled coils in the classical sense. Nonetheless, as judged by SOCKET, we find that the knobs-into-holes interlacing of side chains is maintained locally in the distorted arrangements. This adds a new twist to Crick’s original hypothesis (Crick, 1953), hence our expression extended knobs-into-holes packing. Again, a more-detailed answer to this question, involving geometrical characterisation of the distortions, awaits further analyses of coiled-coil structures.

We envisage that our proposals and the deconvolutions of the helical interfaces that we describe, which fall out of SOCKET analyses, will prove useful to those involved with coiled-coil structure analysis, prediction, engineering and design.

Acknowledgments

We thank members of the Protein Design Group at Sussex for useful discussions and the MRC of the UK for funding.
References
