PROTEINS: Structure, Function, and Genetics 15:223–234 (1993)

Pitch Diversity in α-Helical Coiled Coils

John Seo and Carolyn Cohen
Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02254-9110

ABSTRACT **Two complementary methods for measuring local pitch based on heptad position in α-helical coiled coils are described and applied to six crystal structures. The results reveal a diversity of pitch values: two-stranded coiled coils appear to have pitch values near 150 Å; the values for three- and four-stranded coiled coils range closer to 200 Å. The methods also provide a rapid and sensitive gauge of local coiled-coil conformation. Polar or charged residues in the apolar interface between coiled-coil helices markedly affect local pitch values, suggesting a connection between pitch uniformity and coiled-coil stability. Moreover, the identification of a skip residue (heptad frame shift) in the hemaglutinin glycoprotein of influenza virus (HA) allows interpretation of local pitch changes. These results on relatively short coiled-coil structures have relevance for the much longer fibrous proteins (many of which have skip residues) whose detailed structures are not yet established. We also show that local pitch values from molecular dynamics predictions of the GCN4 leucine zipper are in striking agreement with the high-resolution crystal structure—a result not readily discerned by direct comparison of atomic coordinates. Taken together, these methods reveal specific aspects of coiled-coil structure which may escape detection by global analyses of pitch.**

Key words: leucine zipper, protein folding, molecular dynamics, skip residue, hemaglutinin

INTRODUCTION

The α-helical coiled coil is a widespread and instructive motif in proteins. A good understanding of this fold illuminates the structural principles underlying many different protein classes (for review see refs. 1, 2).

The key feature of the coiled-coil design is the seam of interlocking hydrophobic residues between α-helical chains which stabilize the structure. These residues occur in positions *a* and *d* of the so-called heptad repeat in the amino acid sequence of each chain and are, on average, 3.5 residues apart along the seam. Now the number of residues per turn in a right-handed α-helix has been shown to vary between 3.50 and 3.65 in globular proteins. Correspondingly, there is an apolar stripe formed by residues *a* and *d* that winds round the surface of the α-helix. To maintain side chain meshing when two (or more) helices interlock, the axis of the α-helix takes on a left-handed twist so that the number of residues per turn in the coiled coil is reduced to 3.5.

The pitch or axial repeat per turn of the coiled coil depends critically on the small difference between the 3.5 residue period and the number of residues per turn in the α-helical chains before supercoiling. The pitch is an essential parameter for coiled coils in fibrous proteins because the recognition sites for intermolecular interactions will be positioned depending on the exact pitch. Note that a very small change (say 1%) in the number of residues per turn in the α-helix produces a large change (say 25%) in pitch (see below).

The X-ray fiber diagram of a hydrated molluscan muscle with large amounts of the coiled-coil protein paramyosin first showed the diagnostic near-equatorial layer line, whose spacing gives the ratio of pitch to number of strands in the structure.[3] On this basis, the first estimate made for the pitch was about 180 Å, precisely that proposed by Crick[4] assuming undistorted α-helices with 3.60 residues/turn. This early X-ray work suffered, however, from the quality of the fiber diagrams. Subsequently, estimates of pitch based on fiber diagrams varied from 140 to 170 Å for (two-stranded) α-keratin,[5] 140 Å for (four-stranded) honey bee silk[6] and 136 to 140 Å for paramyosin.[7,8] Coiled-coil pitch was first determined from a crystal structure in the case of tropomyosin at 15 Å,[9,10] and more recently at 9 Å.[11] Here the pitch was found to be about 140 Å. This result also confirmed the prediction of a pitch of 137 Å for tropomyosin made on theoretical grounds.[12,13] Electron microscopy of crystalline sheets of a portion of myosin rod[14] and oothecal protein[15] gave a similar figure. These recent results thus indicate that α-fibrous proteins all have a pitch of about 140 Å, implying that the number of residues per turn in their undistorted α-helices is about 3.636.

Received July 23, 1992; revision accepted November 5, 1992.
Address reprint requests to Dr. Carolyn Cohen, Rosenstiel Basic Medical Science Research Center, Brandeis University, Waltham, MA 02254-9110.

The first atomic level crystal structure determination of a canonical α-helical coiled coil has recently been carried out on the short (33 residue) leucine zipper portion of the yeast transcription factor GCN4.[16] The value for the pitch of this structure was determined to be 180 Å. Using the GCN4 coordinates, Phillips (personal communication and ref. 17) remeasured this parameter and found it to be 142 Å; Phillips also suggested a simple method for measuring pitch. One of the problems that seemed apparent from the GCN4 work was the difficulty with present methods for determining the pitch of a short segment of coiled coil.

We report here two simple complementary methods for this determination which require no fitting of a helical axis to the individual helices of the coiled coil—a potential source of ambiguity in present methods for measuring pitch. We apply these methods to GCN4, and to a number of other coiled-coil-like motifs whose atomic structures have been determined: these include catabolite gene activator protein (CAP) and the GAL4 yeast transcription factor (two-chain parallel), seryl-tRNA synthetase (SRSEC) (two-chain antiparallel), hemagglutinin (HA) (three-chain parallel), and the RNA-binding protein ROP (four-chain antiparallel) (see Fig. 1). The results reveal some of the factors affecting pitch, and the diversity of pitch values in different motifs. Our findings may also be useful for those involved in studies of protein folding and design, and energy minimization techniques.

METHODS

Two complementary methods for determining pitch in a two-chain, parallel coiled coil are described. The Appendix gives the details of the derivation and extends the methods to coiled-coil bundles with more than two chains and antiparallel orientation.

An accurate method of determining local supercoil pitch (here, termed "heptad pitch") may be derived from the coordinates of residue pairs separated by seven amino acids along each chain. Here, we chose to use carbon α positions. In a uniform, canonical coiled coil, individual heptad positions occupy equivalent positions relative to the supercoil axis. For example, the *a* and *d* positions form the core of the dimer interface and the *b*, *c*, and *f* positions occupy exposed positions on the coiled-coil surface. A consequence of this structural equivalence is that the pitch of a supercoil is reflected in the heptad pitches.

Two complementary ways of obtaining a heptad pitch are illustrated in Figure 2. Using two carbon α coordinates from each helix to define local helical paths, the first method (Fig. 2A) obtains a heptad pitch from interhelical separation ($2R$) and crossing angle (θ_{AB}); the second method (Fig. 2B), from axial rise (D) and twist (θ_{12}). The second method was derived independently by Phillips.[17] The resulting pitch expressions (see the Appendix) are:

$$P_{AB} \simeq 2\pi[R/\tan(\theta_{AB})], \quad (1)$$

$$P_{12} = 2\pi[D/\theta_{12}]. \quad (2)$$

Note that (1) and (2) are of the general form 2π(distance/twist). A simple interpretation of this result is that local pitch is estimated by extrapolating a local distance, with an accompanying twist, to a full turn (2π) of the supercoil. Note also that (1) is an approximation, accurate to about 0.5% for typical pitch values; in this report, we actually use a correction term on (1) to increase accuracy to 0.01% (see the Appendix).

In a perfect coiled-coil structure, both (1) and (2) will yield the same heptad pitch; however, if nonuniformities exist—especially those which cause asymmetries about the supercoil axis—the two methods will yield different answers. This difference occurs because the helices are being viewed from orthogonal directions: (1) measures distance and twist perpendicular to the supercoil axis (Fig. 2A), while (2) measures these parameters along the supercoil axis (Fig. 2B). Thus one method emphasizes perturbations in interhelical separation, whereas the other is more sensitive to changes in rise along the supercoil axis. Because each view is complementary to the other, we take the average result of the two methods and use any divergences as an indicator of distortions in the coiled-coil fold (see below). The parallel use of (1) and (2) generally increases both the sensitivity and reliability of pitch calculations.

For a given coiled-coil amino acid sequence, heptad pitches are plotted starting from the N-terminal residue. This convention leads to a "pitch profile" for the first n–7 N-terminal residues. The pitch profile is useful for revealing general trends in local pitch. Heptad pitch values are also labeled by sequence and heptad position in order to emphasize their distinct structural origins in the coiled-coil structure: *a* and *d* heptad pitches apply to amino acids with side chains in the "core" of the coiled-coil's hydrophobic interface; *e* and *g* heptad pitches, the "outer core" positions of the hydrophobic interface; and *b*, *c*, and *f* heptad pitches, the "coat" positions located on the surface of the coiled coil.

In order to determine the correct heptad position of a residue, the interhelical separation distances were examined. Interhelical separations show a distinctive pattern for canonical coiled coils, with the core positions *a* and *d* having the least interhelical separation and the coat position *f* having the most extreme separation. Any rotation of the helices relative to the coiled-coil axis can also be detected by examining interhelical separations for all residues. If helical rotation is present, *b* and *c* as well as *a* and *d* separations become unequal.

Anomolous heptad pitch values frequently arise at isolated points in a structure. This raises the question of what residues—especially ones at the termini—should not be considered part of the coiled-coil fold. An anomolous heptad pitch value alone is not a satisfactory criterion; neither is the type of residue in a given heptad position. Residues were omitted from our analysis only if pitch values from (1) and (2) diverged by more than 20% in value. Either (1) or (2) alone will always yield a pitch value, regardless of the conformation of the protein fold; however, agreement between the two methods ensures that the fold is interpretable as being coiled coil from at least two different points of view. We have found in practice that the largest divergences between methods 1 and 2 occur at the ends of coiled coils, so that this divergence is useful to delineate coiled-coil boundaries.

RESULTS

GCN4

The coiled coil of the GCN4 leucine zipper is the dimerization domain of this yeast transcription factor. This two-stranded, parallel coiled coil is approximately four heptads long and is stable when isolated from its adjoining DNA-binding region. GCN4 provides the first high resolution structure of a canonical coiled coil.[16] The amino acid composition of GCN4, and leucine zippers in general, is similar to that of tropomyosin and myosin rod—having roughly 40% charged, 35% apolar, and 20% polar residues. There are more acidic than basic residues in fibrous proteins, but this is not generally true in leucine zippers. The pitch profile of GCN4 is shown in Figure 3A.

As seen from Figure 3A, the heptad pitch is fairly uniform along the length of the GCN4 coiled coil, averaging about 147 Å, but variations exist. The heptad pitch of the coat residues starts off at the N-terminus with a value around 148 Å, rises to 158 Å in the middle section, and then gradually comes back down to 144 Å at the C-terminus. Correspondingly, the twist of the coiled coil becomes slightly flattened out in the middle.

The significant disruptions in heptad pitch appear to occur exclusively in the core position *a*. The largest perturbation occurs about a third of the way from the N-terminus due to the presence of the polar side chain of Asn-16 in the hydrophobic core. The large rise in pitch caused by Asn-16 has relatively little effect on adjacent heptad pitches, indicating isolation of effects between heptad pitches. A related observation is that the heptad pitches of the leucines in the *d* position—which are conserved in the dimer interfaces of most leucine zippers—are correspondingly uniform. This result suggests that even when surrounding side chains vary, the large, apolar leucine side chains are stably anchored in the hydrophobic core.

CAP

The coiled-coil region of the catabolite gene activator protein (CAP) plays a central role in the dimerization of this DNA-binding protein. The surface of the CAP coiled coil is not solvent exposed, but is flanked on both sides by globular domains which bind cyclic AMP and DNA.[18] The two-chained, parallel CAP coiled coil is approximately three heptads in length. Relative to GCN4, the CAP coiled coil has a higher apolar content and an unusually low number of charged residues; both of these features are consistent with the buried position of the coiled coil in the CAP dimer.

The CAP pitch profile (Fig. 3B) is similar to that of GCN4, but the overall pitch of the CAP coiled coil is lower, averaging about 132 Å. Disruptions in heptad pitch again occur exclusively in the core position *a* with the other core position *d* exhibiting a very stable pitch value around 136 Å. The largest perturbation in heptad pitch is caused by Ser-117 which occurs a third of the way from the N-terminus of the coiled coil at an *a* position. Note that here, the presence of the polar Ser-117 side chain in the dimer interface causes a depression of heptad pitch rather than an increase as with Asn-16 of GCN4.

GAL4

A short, parallel coiled-coil region spanning only two heptads has been shown to facilitate dimerization in an N-terminal GAL4 fragment bound to DNA.[19] This short coiled coil does not promote dimerization unless GAL4 is bound to DNA. In the complete protein, the GAL4 coiled coil dimerization element continues for another heptad beyond the two of the GAL4 fragment so that its full size is similar to that of CAP. The amino acid content of the shortened dimerization element is typical of that found in GCN4. The GAL4 pitch is consistently low (Fig. 3C), averaging around 93 Å, indicating that the helix crossing angle is increased roughly 20° above that of CAP.

The smooth GAL4 pitch profile is consistent with the fact that no polar or charged residues have invaded the core positions. The N-terminal residue of the coiled coil has been omitted from the analysis because of a 50% divergence in pitch values, using methods (1) and (2). The GAL4 apolar core—composed of three leucines and one valine—does, however, contain an important singularity. The lone valine of the core occurs in the middle of the segment and causes the dimer interface of each helix to be locally concave (due to the reduced side chain volume of valine relative to leucine). This local concavity increases the helix crossing angle in order to maximize hydrophobic contact between helices. It is possible that if the helices were longer than two heptads, changes in pitch due to the GAL4 valine would have had a more localized effect in the core.

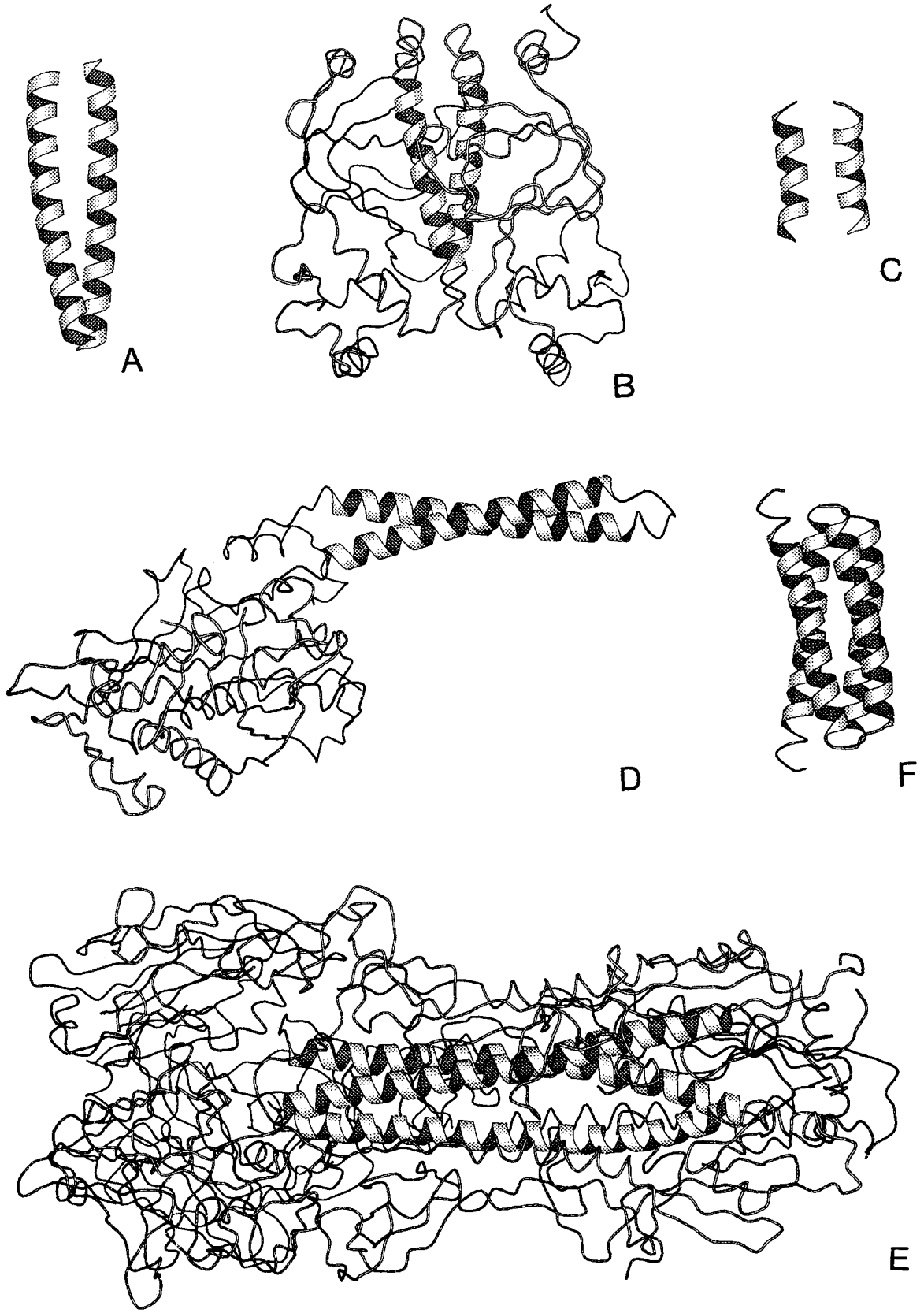

Fig. 1.

SRSEC

The two-stranded, antiparallel coiled-coil arm of seryl tRNA synthetase (SRSEC) projects out from each globular domain of the protein dimer into solution.[20] The SRSEC arm binds RNA and is just over four heptads long (only two residues longer than GCN4). Its amino acid content is similar to that of GCN4. Because of the antiparallel orientation of the helices, the heptad repeats must be staggered, rather than side by side, to keep the hydrophobic residues in register. Despite the helices' antiparallel orientation, the overall structure of the SRSEC coiled coil is similar to that of GCN4. In particular, extensive hydrophobic contacts are made in the core by leucine–(iso)leucine pairs. The distal half of the arm does, however, show significant invasion of the core by charged and polar residues.

The SRSEC pitch profile (Fig. 3D) shows that the pitch is relatively stable around a value of 150 Å in the first, proximal half of the arm, but starts to increase half-way through the arm, reaching 250 Å at the distal end. The rise in pitch seen in the distal half of the arm correlates well with disruption of the hydrophobic core by charged and polar residues. The flattening out of superhelical twist in the distal half of the arm, in conjunction with the disrupted core of this region, might facilitate bending of the arm upon binding to RNA.

Fig. 1. Ribbon diagrams of six crystal structures with coiled-coil motifs (to scale). Coiled-coil regions are shown as ribbons; noncoiled-coil regions, as strands. (Note that here isolated helices are also illustrated as strands.) Brookhaven Protein Data Bank accession numbers are given in brackets; otherwise, coordinates were obtained from the authors (see text for references); also given in brackets is the resolution of the crystal structure determination. (**A**) GCN4-p1 [1.8 Å]: this peptide structure corresponds to the leucine zipper dimerization domain of the GCN4 yeast transcription factor. In the native dimer, DNA-binding domains are located immediately N-terminal to the coiled-coil region of the figure (towards the top of the page). (**B**) CAP [3GAP, 2.5 Å]: like the GCN4 leucine zipper, the catabolite gene activator protein also binds DNA. The N-terminus of the coiled coil region is directed toward the top of the page. (**C**) GAL4 [2.7 Å]: this coiled-coil fragment forms a dimerization element in a 65-residue N-terminal fragment of the GAL4 yeast transcriptional activator. In the complete GAL4 structure, a DNA-binding domain is located N-terminal to this coiled-coil (toward the top of the page). (**D**) SRSEC [2.5 Å]: the seryl-transfer RNA synthetase from *E. coli* is unique in that it does not contain a classic β–α–β dinucleotide-binding domain to facilitate its binding to tRNA. Preliminary studies indicate that the antiparallel coiled-coil arm which forms an N-terminal domain of this structure binds its cognate tRNA when charging it with serine. Only one monomer of the dimeric enzyme is shown. (**E**) HA [2HMG, 3.0 Å]: the hemagglutinin glycoprotein of influenza virus structure is a trimer, the monomers of which are cleaved into two separate chains HA_1 and HA_2: HA_1 comprises the globular domain which contains a host cell receptor binding site (on the left of the figure); HA_2 makes up the coiled-coil stem which provides the major stabilizing forces for the trimeric subunit interactions (N-terminus of the coiled-coil region is to the left of the figure). (**F**) ROP [1ROP, 1.7 Å]: this protein participates in the control of plasmid replication via regulation of an RNA–RNA interaction. The ROP structure is a dimer whose monomers are almost entirely composed of α-helices. The N-terminus of each monomer is located on the left of the figure. Figure produced using MOLSCRIPT.[35]

HA

The fibrous stem of the hemagglutinin glycoprotein of influenza virus (HA) is composed of a triple-stranded, parallel coiled coil spanning seven heptads.[21] The HA coiled coil positions a globular, receptor binding site approximately 76 Å from the viral membrane. The coiled-coil stem exhibits two distinct regions. The N-terminal region (four heptads), which is closest to the receptor binding site, is of special interest because it contains a "drift" in heptad frame.[21] We show later (see Discussion) that this drift is due to a skip residue (a single amino acid insertion which shifts the heptad repeat pattern). The last three C-terminal heptads comprise the second region which exhibits a complete loss of apolar residues in the core and a divergence of helices from the coiled-coil axis. Although the charge is essentially balanced in the C-terminal region, the skip residue region exhibits a higher proportion of acidic than basic residues.

The pitch profile is periodically disrupted virtually throughout the four heptad N-terminal region (Fig. 3E). The peaks in pitch range in value from 600 to 1600 Å and occur only in core or outer core residues. These same positions show some distortion from the coiled-coil conformation as indicated by divergences in (1) and (2). The coat positions exhibit a rise in heptad pitch in the region of the skip residue, but the pitch values are smaller (from 250 to 350 Å) than in the core position. This persistent pitch differential between the core and coat positions indicates that the individual α-helices are rotating relative to the trimer interface. This rotation is expected in the case of a shift in heptad frame of the amino acid sequence: in response to the frame shift, the helices adjust their orientations relative to the trimer interface; otherwise, apolar side chains within the shifted heptad frame will be solvent exposed, and charged/polar side chains will be buried in the trimer interface. Thus, the coat/core pitch differential of the HA coiled coil illustrates how heptad pitch values may signal a heptad frame shift in the amino acid sequence.

The pitch apparently stabilizes in the last C-terminal three heptads, settling to 155 Å at the C-terminus (Fig. 3E)—a somewhat surprising result in view of the complete loss of apolar residues in the trimer interface of this region. This finding indicates that despite the loss of packing of apolar side chains in the core, the helices of this trimeric coiled coil can still maintain typical pitch values.

ROP

The RNA-binding protein ROP, which is involved in the regulation of replication of ColE1 plasmids, is a very regular, four-helix, antiparallel coiled-coil bundle.[22] The ROP structure is composed of two monomers related by an exact dyad perpendicular to

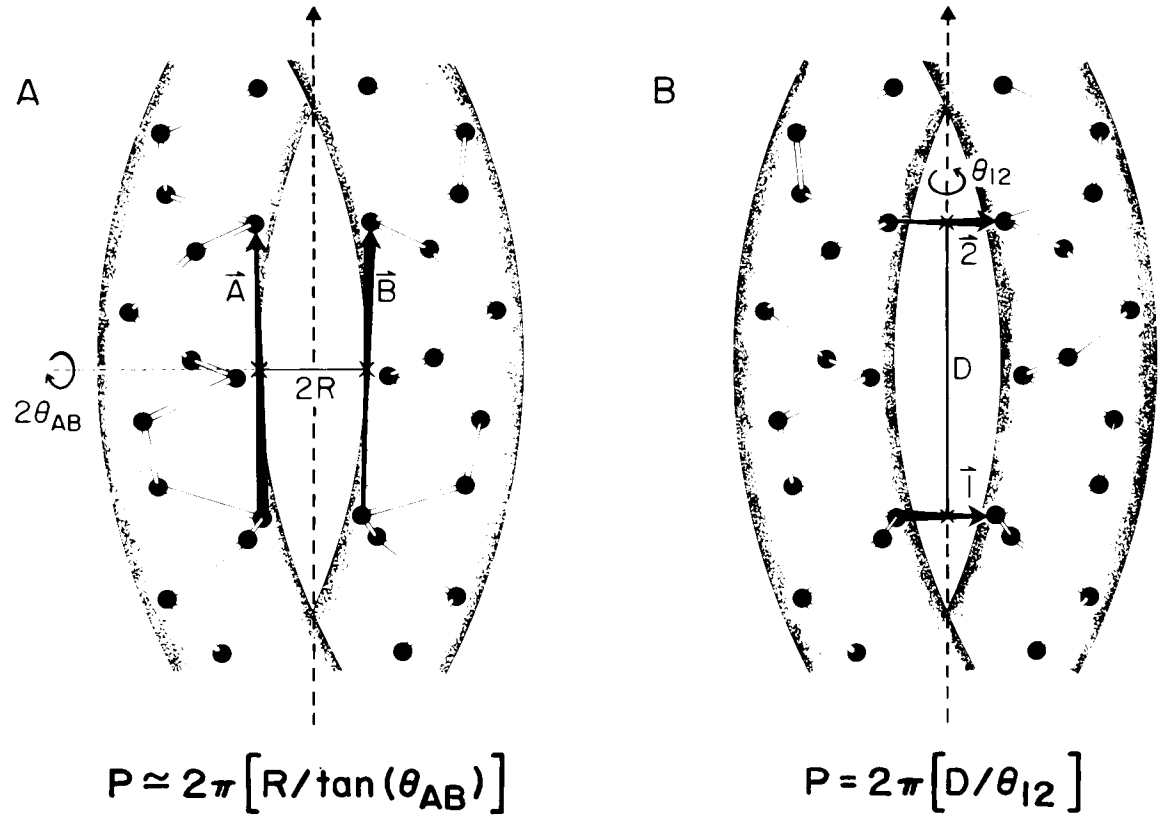

Fig. 2. Two complementary methods for obtaining a local pitch value. Carbon α coordinates of residues spaced seven apart on both helix A and B occupy equivalent positions relative to the coiled-coil axis (vertical, dashed line). Thus, twisting of the helices around the coiled-coil axis may be determined using the coordinates of appropriately spaced residues as helical reference points. The points with crosses (×) denote midpoints between carbon α positions. (**A**) The angle 2 θ_{AB} is taken as the dihedral angle between vectors A and B rotated about the line 2*R*; 2*R* also gives an interhelical separation distance. (**B**) The angle θ_{12} is taken as the dihedral angle between vectors 1 and 2 about the line *D* which also gives an axial rise distance. Note that the local pitch expression in (**A**) is an approximation good to about 0.5%. In practice, we have used a correction term (see the Appendix for the complete expression and extension to coiled-coil bundles).

its long axis. The amino acid content of the ROP coiled coil is similar to that of GCN4. Although each of the ROP helices is four heptads long, because of its symmetry only two unique heptads of interacting residues exist. As in SRSEC, the antiparallel interaction of the ROP helices requires an axial stagger of the heptad repeat in order to bring apolar residues together in the core. The ROP pitch profile (Fig. 3F) indicates that the heptad pitch is relatively uniform in all heptad positions with an average value of 186 Å, consistent with the results of Banner et al.[22] [The end residues of each helix were omitted from the analysis due to a two-fold divergence in (1) and (2).] The small peaks seen near the ends of the pitch profile are correlated with the presence of a glutamine in the core whose polar side chain points toward solvent. The uniformity of the pitch may be related to the high symmetry of the structure.

GCN4 Model Structures

Seven model structures of GCN4 have been generated using an automated algorithm in conjunction with techniques of simulated annealing and molecular dynamics.[23] This family of model structures results from using different initial helix rotations relative to the dimer interface. These models were generated before publication of the X-ray structure of GCN4 and provide a test of the algorithm's predictive ability. The pitch profiles of the model structures are somewhat varied but correlate rather well with the pitch profile of the X-ray structure. The most successful model is illustrated in Figure 4.

DISCUSSION

Overview

The simple heptad pitch methods developed and applied here provide a useful means to analyze and compare coiled-coil structures. With the exception of Phillips,[17] previous methods to determine pitch have been based largely on graphical superposition or least-squares fitting which yield only a global measure of pitch. Such methods are most useful for rather uniform or idealized structures. In the case of real protein structures, however, local features are

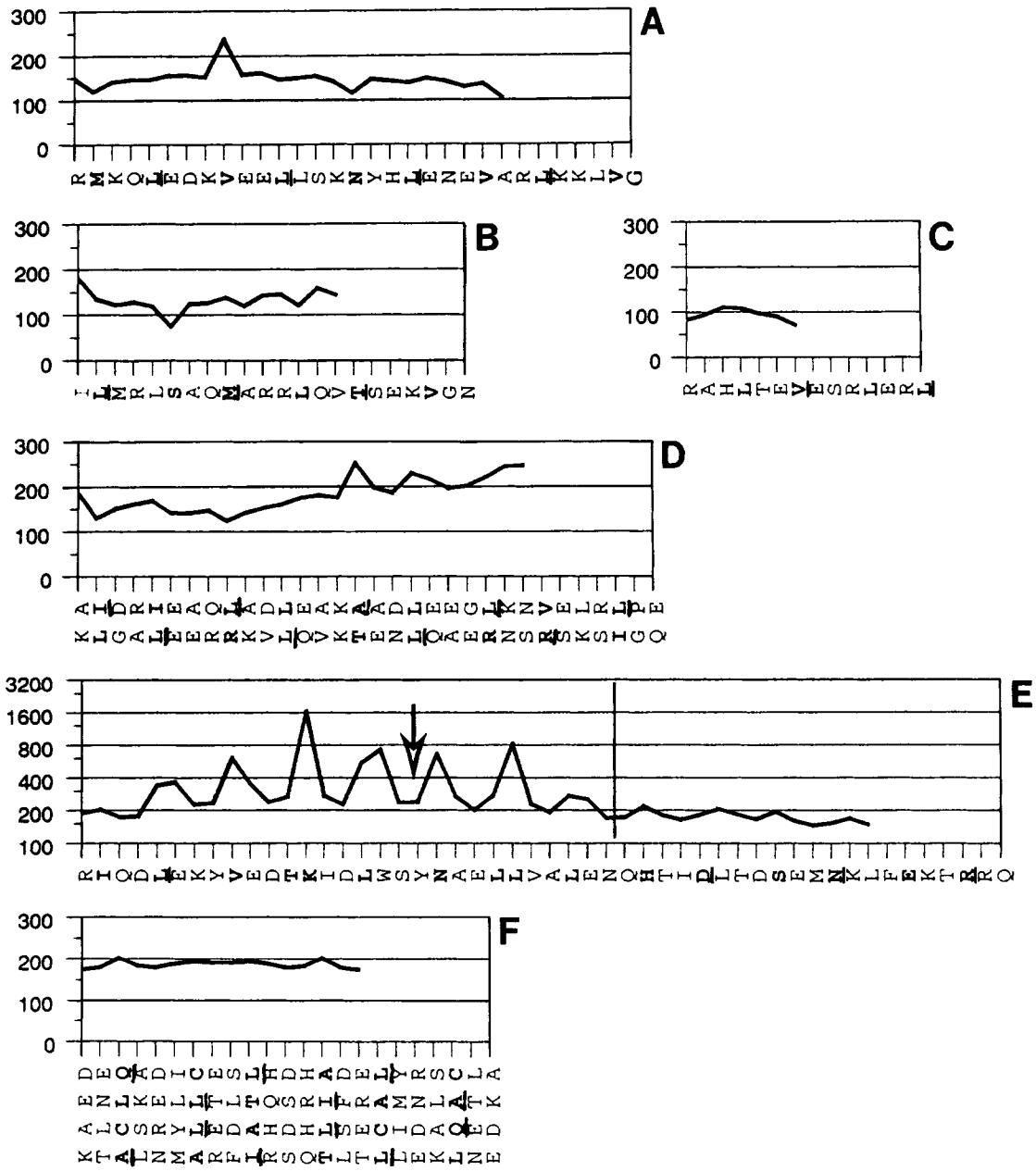

Fig. 3. Pitch profiles for six coiled-coil crystal structures. Pitch values are the average result of methods 1 and 2. As a convention, the heptad pitch value for the n to $n + 7$ residue is plotted at the nth residue. Bold facing signifies a core position in the coiled-coil structure. The d heptad position is also underlined. This identification is not possible in the skip residue region of HA where the structural heptad frame shows a drift. For antiparallel coiled coils, the sequence of all chains is given such that all residues are in axial register. (**A**) GCN4-p1: residues 1–31. (**B**) CAP: residues 112–133. (**C**) GAL4: residues 51–64. (**D**) SRSEC: residues 29–60 (bottom row), residues 100–69 (top row). (**E**) HA: residues 76–125. [Note that here the vertical axis is logarithmic. The arrow marks the tentative skip residue Tyr-94, and the vertical line divides the coiled coil into two regions with distinct structural features (see text for description).] (**F**) ROP: residues 6–28, 54–32, 28–6, 32–54 (from bottom to top row).

often informative. The complementary use of the two heptad pitch methods described here allows rapid inspection of only two residue positions per helix to provide a sensitive gauge of local coiled-coil conformation. Moreover, divergences between these two methods provide a simple means to detect distortions in this protein fold. Taken together, the methods reveal aspects of coiled-coil structure which may escape detection by global analyses of pitch.

Pitch Diversity

The six crystal structures examined here reveal a diversity in heptad pitch values (Fig. 5). Two-stranded coiled coils seem to have lower pitch values

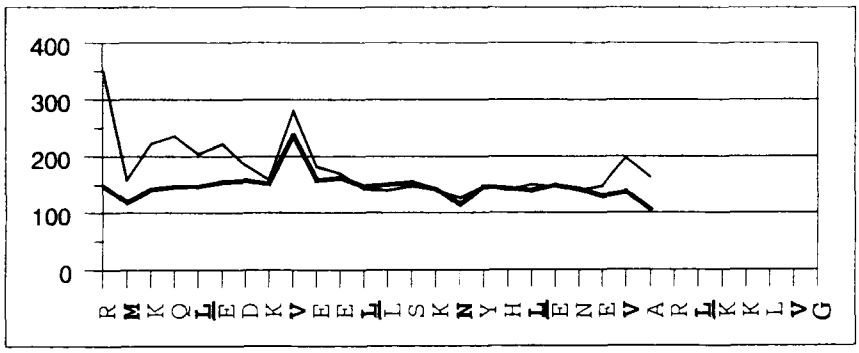

Fig. 4. Comparison between the pitch profiles of the GCN4 crystal structure (heavy line) and a model prediction structure (thin line) generated by Nilges and Brünger.[23] The entire family of model structures exhibits similar agreement with the crystal structure. At least one of the distinctive drops in pitch plotted at Met-2 and Asn-16 or the large peak in pitch plotted at Val-9 are present in all seven model structures, and all drops/peaks are simultaneously present in four of the model structures. Pitch values are the average result of methods 1 and 2.

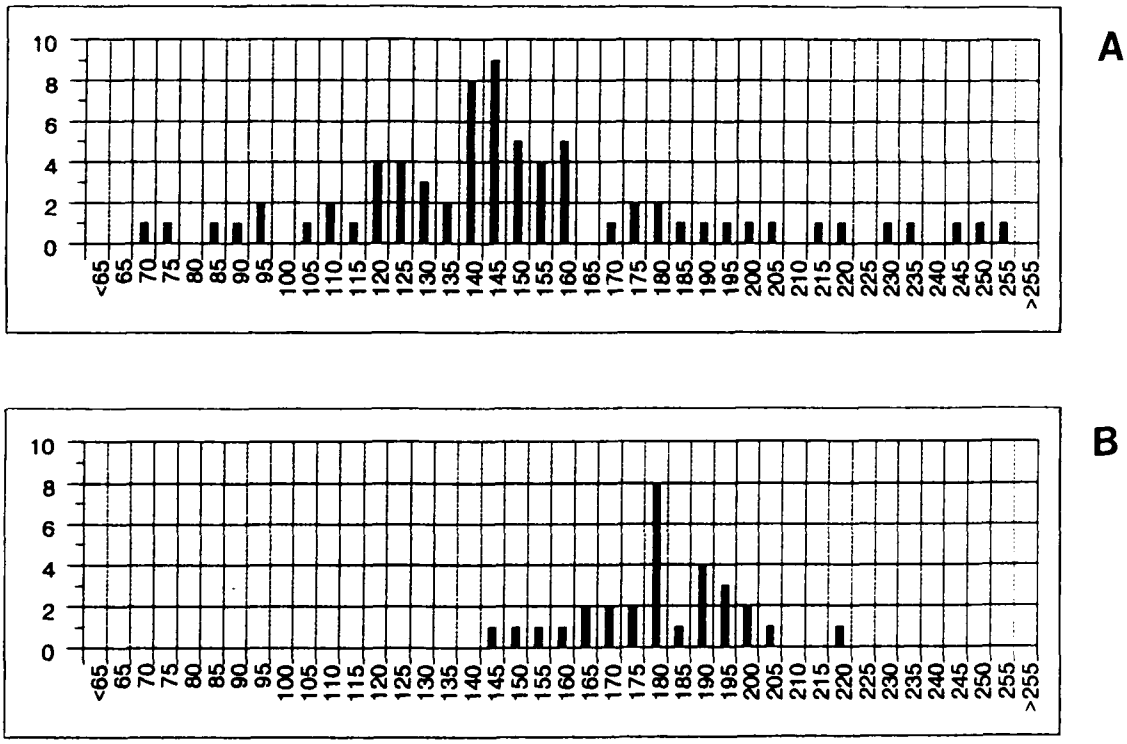

Fig. 5. Histogram of heptad pitch values for six coiled-coil crystal structures. (**A**) Two-stranded coiled coils GCN4, CAP, and GAL4. Note that 7 of the 9 scores below 115 Å are due to GAL4. Total scores: 71. (**B**) Three- and four-stranded coiled coils HA (excluding the skip residue region) and ROP. Total scores: 30. Note that the diversity in local pitch values involves all heptad positions. Pitch values are the average result of methods 1 and 2.

(ranging near 150 Å) than the three- and four-stranded HA and ROP structures (range closer to 200 Å). Correspondingly, the number of residues per turn in the undistorted α-helices decreases from 3.64 to 3.60 as the number of strands increases. This change may be due to the decreased solvent exposure of hydrogen bonds in the α-helices that accompanies an increase in number of strands.[24] Thus, two-stranded structures have the narrowest azimuthal band of apolar residues (hence hydrogen bond shortening and the shortest pitch lengths), whereas three- or four-stranded structures have increasingly wide azimuthal bands with less differential hydrogen bond shortening and consequently greater pitch lengths [see also (1)].

Disruptions in the pitch profile occur almost exclusively in the *a* and *d* (core) heptad positions, indicating that the core has a major influence in mod-

ulating local pitch values. The constant pitch of conserved leucines in the GCN4 core, the association of pitch uniformity with the presence of a leucine–(iso)leucine core in SRSEC, and the general correlation of perturbations in heptad pitch with the presence of polar or charged residues in the core all suggest a connection between coiled-coil stability and local pitch uniformity.[25]

The pitch analysis of uniform sections of coiled coils with singular residues provides an opportunity to examine the effects of a perturbation on the surrounding protein fold. Strong isolation effects are present in the structures examined: for example, both the perturbation in pitch caused by Asn-16 in GCN4, Ser-117 in CAP, Gln-34 in ROP, and the pitch differential of the HA "skip residue" region show that most distortions in the core have only a slight effect on the heptad pitch of coat residues. This finding implies that coiled coils are an ideal system for protein engineering studies[26] whereby a "guest position"[27] is used to test the effect of single amino acid substitutions. Note that the results on GAL4 suggest that a coiled coil must be more than two heptads long to prevent local packing distortions from affecting the entire fold.

Fibrous Proteins: Global vs. Local Pitch

The finding that relatively short two-stranded coiled coils have local pitch values around 150 Å is consistent with the pitch range of 140±10 Å determined for tropomyosin[10,11] and for other two-stranded fibrous proteins (see Introduction). In these proteins too, however, it is likely that local pitch values vary significantly in regions where there is a distortion of the dimer interface produced by the presence of polar and charged residues or by poor apolar packing contacts, or where skips or stutters occur (see also below). The location of such local pitch variations is much more readily detected by methods of sequence analysis than by low resolution X-ray crystallographic structure determinations. In certain fibrous muscle proteins, such as tropomyosin, myosin rod, and paramyosin, such regions appear to be relatively rare along the long coiled-coil rods.[1,28]

The cumulative pitch over multiple heptads (here termed "global pitch") is not the average of local pitch values. Because the axial rise per heptad is approximately constant, local variations in pitch are produced mainly by variations in axial twist which is inversely proportional to pitch. A good estimate of global pitch is therefore the inverse of the average of inverse local pitch values. (The nature of such an inverse average is that the global pitch will always be less than the average of local pitch values. This point is illustrated by considering the effect of including one heptad with infinite pitch (zero twist): the average of the local pitch values would be infinite, whereas the global pitch would merely reflect the zero twist in a single section.) It should be noted that method (2) used by Phillips[17] for GCN4 provides a simple way to measure the global pitch of a structure starting from an N-terminal heptad position to an equivalent heptad position at the C-terminus. This approach, however, is sensitive to local perturbations at the ends of the coiled coil and should therefore be used only after assessing the pitch profile of the structure.

The relative constancy of global pitch for certain fibrous proteins is important for the assembly of these structures. This is because the recognition sites on these proteins will be positioned depending on the exact pitch of the coiled coil. Thus, in the case of tropomyosin, the (chiefly) coat residues of each of the seven half turns of the coiled coil on the molecule forms a recognition site which interacts with an actin monomer in the thin filament. Similarly, the coat residues of myosin rod and paramyosin display alternate zones of opposite charge so that the assembly and coassembly of these molecules appear to be determined by complementary charge interactions. It appears that the global pitch of ~140 Å for these muscle proteins derives both from their two-strandedness, and from their general conservation of sequences, especially in the core positions. In the case of shorter structures such as leucine zippers, which do not have multiple cooperative binding sites, local pitch variations may play critical roles in affecting molecular recognition.[29]

Skip Residues and Heptad Frame Shifts

The structure of the HA skip region may have relevance for similar regions containing skips and stutters in fibrous proteins and other shorter coiled coils. In the case of fibrous proteins where there is striking continuity in the heptad repeat patterns, additional skip residues, which cause a shift in heptad frame, are easy to detect by inspection. In short coiled coils, however, such frame shifts may be overlooked. Lupas et al.[2] have recently developed a method which identifies the "best" heptad frame for a given sequence. When used with a 28-residue sliding window, this method detects skips and stutters previously identified in a variety of fibrous proteins. In HA, the presence of a skip region was not readily seen by sequence inspection.[30] Application of the Lupas et al. method, however, using a sliding window 7 to 11 residues wide, indicates that a heptad frame shift occurs at or near Tyr-94 (which occurs near the middle of the N-terminal region) consistent with the HA pitch results (Fig. 3E). We note that a sliding window wider than 13 residues failed to detect the HA skip residue, suggesting that the 28-residue window originally used by Lupas et al.[2] may be too big to resolve certain details in some coiled-coil sequences.

Previously, no X-ray crystal structure had been analyzed to establish the effect of a skip on the pa-

rameters of a coiled coil. A careful model-building study was carried out, however, by McLachlan and Karn[31] on the four skip regions of the nematode myosin rod. Their results indicated that the hydrophobic seam shifted gradually across the helix surface over two to three heptads, producing an increase in local pitch, and that there were no sharp breaks or distortions in the coiled coil. The N-terminal region of the 3 Å HA crystal structure provides good confirmation of these model predictions, showing a clear drift in heptad frame over roughly three heptads[21] and a local increase in overall pitch (Fig. 3E). The atomic structure reveals, however, that the increase in overall pitch of the skip region is rather less than previously proposed. Although HA is a somewhat unusual coiled-coil structure, these results imply that the functional role of a skip residue may be related not only to modulation of pitch, but also to local coil-coil destabilization. Offer[32] has, in fact, recently correlated skip regions with local bends in the myosin rod.

Molecular Dynamics Methods

The striking agreement between the pitch profiles of a GCN4 model[23] and the corresponding crystal structure (Fig. 4) indicates that successful modeling of some key structural features of coiled coils is within the reach of present molecular dynamics methods. Note that previous evaluations of GCN4 model structures,[33] using standard statistical measures of comparison with the crystal structure, do not reveal detailed features of similarity as successfully as the heptad pitch methods applied here. Differences between the models and the X-ray structure (data not shown) are concentrated around the main peak in the X-ray pitch profile, with the model heptad pitch values almost always rising above the X-ray values. This result may be a consequence of the fact that the models are actually solution structures, i.e., they do not take into account lattice constraints in the GCN4 crystal structure. The major peak in the GCN4 pitch profile, due to the presence of Asn16 in the dimer interface (Fig. 3A), is related to the "asymmetric" arrangement of the side chains of this pair of residues about the coiled-coil axis in the crystal structure. Because NMR studies indicate that Asn-16 is symmetric in solution, it has been suggested[16] that the GCN4 crystal structure is trapped at Asn-16 in one of a number of conformations energetically available to the solution structure. The variation of heptad pitch values of models around the disruption caused by Asn-16 may be a reflection of the multiple conformational states accessible to GCN4 in solution. Thus, the range of model structures may, in fact, correspond well with a family of NMR solution structures. If this is so, the various model conformations may provide information on the dynamic behavior of coiled coils in solution.

Protein Folding

Coiled coils have special significance for the protein folding problem since the heptad repeat allows the recognition of tertiary structure from the primary sequence alone[4] (see also ref. 1). Moreover, the recent development of statistical measures by Lupas et al.[2] has provided objective methods for detecting heptad frame shifts and breaks in coiled coils from such sequence data. The large data base of sequences for coiled-coil structures allows a sound statistical examination of the tendencies of different amino acids to occur in different heptad positions. We note that these "heptad statistics" as obtained by Lupas et al.[2] essentially correspond to the "3D-1D" scores used by Bowie et al.[34] in the construction of protein structure profiles. In fact, the method of Lupas et al. is analogous to a special application (for coiled coils) of the more general method of Bowie et al. for globular proteins. This connection can be made by recognizing the correspondence between heptad positions in the core (*a*, *d*), outer core (*e*, *g*), and coat (*b*, *c*, *f*) and the α-helical buried ($B_{1\alpha}$, $B_{2\alpha}$, $B_{3\alpha}$), partially buried ($P_{1\alpha}$, $P_{2\alpha}$), and exposed (E_{α}) side chain environment categories of Bowie et al.,[34] respectively. For example, according to Bowie et al., leucine is found 3.6 times more often than would be expected at random in an apolar-buried helix position in globular proteins ($B_{1\alpha}$); in coiled coils, Lupas et al. find that the corresponding number is 3.9 in the *d* position. Similar correlations are found for the majority of the amino acids. This observation demonstrates the direct relevance of coiled-coil studies to the general understanding of both fibrous and globular protein structures.

The heptad pitch methods developed here provide a means for relating native, engineered, or even modeled coiled-coil structures to heptad sequence studies. By correlating modulations in coiled-coil structure with the presence of specific residues, heptad pitch analyses can provide a useful quantitative tool for assessing the individual roles of amino acids in protein structures. This information represents one essential element in the solution of the protein-folding problem.

ACKNOWLEDGMENTS

We are grateful to David A.D. Parry for discussion and criticism of this paper. We are also grateful to George N. Phillips, Jr. for providing results prior to publication. This work was supported by grants to CC from the National Institutes of Health (AR17346), National Science Foundation (DMB90-04746), and the Muscular Dystrophy Association. John Seo was a Postdoctoral Fellow of the Muscular Dystrophy Association.

REFERENCES

1. Cohen, C., Parry, D.A.D. α-Helical coiled coils and bundles: How to design an α-helical protein. Proteins 7:1–15, 1990.

2. Lupas, A., Van Dyke, M., Stock, J. Predicting coiled coils from protein sequences. Science 252:1162–1164, 1991.
3. Cohen, C., Holmes, K.C. X-ray diffraction evidence for α-helical coiled coils in native muscle. J. Mol. Biol. 6:423–432, 1963.
4. Crick, F.H.C. The packing of α-helices: Simple coiled-coils. Acta Crystallogr. 6:689–697, 1953.
5. Fraser, R.D.B., MacRae, T.P., Miller, A. The coiled-coil model of α-keratin structure. J. Mol. Biol. 10:147–156, 1964.
6. Atkins, E.D.T. A four-strand coiled-coil model for some insect fibrous proteins. J. Mol. Biol. 24:139–141, 1967.
7. Elliott, A., Lowy, J., Parry, D.A.D., Vibert, P.J. Puzzle of the coiled coils in the α-protein paramyosin. Nature (London) 218:656–659, 1968.
8. Elliott, A., Lowy, J. A model for the coarse structure of paramyosin filaments. J. Mol. Biol. 53:181–203, 1970.
9. Phillips, G.N., Jr., Lattman, E.E., Cummins, P., Lee, K.Y., Cohen, C. Crystal structure and molecular interactions of tropomyosin. Nature (London) 278:413–417, 1979.
10. Phillips, G.N., Jr., Fillers, J.P., Cohen, C. Tropomyosin crystal structure and muscle regulation. J. Mol. Biol. 192: 111–131, 1986.
11. Whitby, F.G., Kent, H., Stewart, F., Stewart, M., Xie, X., Hatch, V., Cohen, C., Phillips, G.N., Jr. Structure of tropomyosin at 9 Angstroms Resolution. J. Mol. Biol., 227:441–452, 1992.
12. Parry, D.A.D. Analysis of the primary structure of α-tropomyosin. J. Mol. Biol. 98:519–535, 1975.
13. McLachlan, A.D., Stewart, M. Tropomyosin coiled-coil interactions: Evidence for an unstaggered structure. J. Mol. Biol. 98:293–304, 1975.
14. Quinlan, R.A., Stewart, M. Crystalline tubes of myosin subfragment-2 showing the coiled-coil and molecular interaction geometry. J. Cell Biol. 105:403–415, 1987.
15. Bullough, P.A., Tulloch, P.A. High-resolution spot-scan electron microscopy of microcrystals of an α-helical coiled-coil protein. J. Mol. Biol. 215:161–173, 1990.
16. O'Shea, E.K., Klemm, J.D., Kim, P.S., Alber, T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science 254:539–544, 1991.
17. Phillips, G.N., Jr. What is the pitch of the α-helical coiled coil? Proteins: 14:425–429, 1992.
18. Weber, I.T., Steitz, T.A. Structure of a complex of catabolite gene activator protein and cyclic/amp refined at 2.5 angstroms resolution, J. Mol. Biol. 198:311–326, 1987.
19. Marmorstein, R., Carey, M., Ptashne, M., Harrison, S.C. DNA recognition by GAL4: Structure of a protein-DNA complex. Nature (London) 356:408–414, 1992.
20. Cusack, S., Berthet-Colominas, C., Hartlein, M., Nassar, N., Leberman, R. A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. Nature (London) 347:249–255, 1990.
21. Wilson, I.A., Skehel, J.J., Wiley, D.C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution, Nature (London) 289:366–373, 1981.
22. Banner, D.W., Kokkinidis, M., Tsernoglou, D. Structure of the ColE1 Rop Protein at 1.7 Å Resolution. J. Mol. Biol. 196:657–675, 1987.
23. Nilges, M., Brünger, A.T. Automated modeling of coiled coils: Application to the GCN4 dimerization region. Protein Eng. 4:649–659, 1991.
24. Goodman, E.M., Kim, P.S. Periodicity of amide proton exchange rates in a coiled-coil leucine zipper peptide. Biochemistry 30:11615–11620, 1991.
25. Alber, T. Structure of the leucine zipper. Curr. Opin. Genet. Dev. 2:205–210, 1992.
26. Hodges, R.S. Unzipping the secrets of coiled-coils. Curr. Biol. 2:122–124, 1992.
27. O'Neil, K.T., DeGrado, W.F. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. Science 250:646–651, 1990.
28. Dibb, N.J., Maruyama, I.N., Krause, M., Karn, J. Sequence analysis of the complete *Caenorhabditis elegans* myosin heavy chain gene family. J. Mol. Biol. 205:603–613, 1989.
29. Lamb, P., McKnight, S.L. Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerization. Tr. Biochem. Sci. 16:417–422, 1991.
30. Ward, C.W., Dopheide, T.A. Influenza virus haemagglutinin. Structural predictions suggest that the fibrillar appearance is due to the presence of a coiled-coil. Aust. J. Biol. Sci. 33:449–455, 1980.
31. McLachlan, A.D., Karn, J. Periodic features in the amino acid sequence of nematode myosin rod. J. Mol. Biol. 164: 605–626, 1983.
32. Offer, G. Skip residues correlate with bends in the myosin tail. J. Mol. Biol. 216:213–218, 1990.
33. Nilges, M., Brünger, A.T. Successful prediction of the coiled coil geometry of the GCN4 leucine zipper domain by simulated annealing; comparison to the X-ray structure. Proteins 15:133–146, 1993.
34. Bowie, J.U., Luthy, R., Eisenberg, D. A method to identify protein sequences that fold into a known three-dimensional structure. Science 253:164–170, 1991.
35. Kraulis, P.J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. 24:946–950, 1991.

APPENDIX

Given the coordinates of a coiled-coil's α-carbon positions (A_n, B_n, . . . , where A_n and B_n are the cartesian coordinates of the nth residue of helix A and B, respectively; for now, we use a two-chain, parallel coiled coil to fix ideas), we may calculate the local superhelical pitch from as few as two pairs of coordinates: A_n, B_n, A_{n+7}, B_{n+7}. Two complementary approaches to calculating the heptad pitch may be used:

$$P_1 = P_{\text{approx}}[1 - 1/6(\pi D/P_{\text{approx}})^2]$$

where

$$P_{\text{approx}} = 2\pi|R|/\tan(\theta_{AB}) \quad \text{(A1)}$$

and

$$P_2 = 2\pi|D|/\theta_{12} \quad \text{(A2)}$$

where

$$2\,\theta_{AB} = A\cos[F_A \cdot F_B/(|F_A|\,|F_B|)]$$

$$\theta_{12} = A\cos[F_1 \cdot F_2/(|F_1|\,|F_2|)]$$

$$2R = (I_1 + I_2)/2$$

$$D = (I_A + I_B)/2$$

$$F_A = I_A \times R \times I_A$$

$$F_B = I_B \times R \times I_B$$

$$F_1 = I_1 \times D \times I_1$$

$$F_2 = I_2 \times D \times I_2$$

$$I_A = A_{n+7} - A_n$$

$$I_B = B_{n+7} - B_n$$

$$I_1 = B_n - A_n$$

$$I_2 = B_{n+7} - A_{n+7}$$

where $A\cos(\,)$ is the inverse cosine function, $F \cdot G$ denotes the dot product of vectors F and G; $|F|$ is the magnitude of vector F; and $F \times G$ is the cross product of F and G. Note that θ angles are taken in a manner analogous to that used to obtain a peptide bond's dihedral angle.

The first expression for pitch (A1) is based upon a

formula for the pitch of a supercoil with helix crossing angle 2θ and helical separation $2R$ [let s be the helical path length over one full superhelical turn; then $P = s \times \cos(\theta)$, and $2\pi R = s \times \sin(\theta)$; combining the two expressions to eliminate s yields $P = 2\pi R/\tan(\theta)$]. The helix crossing angle, $2\theta_{AB}$, is only an approximation because it is taken from the straight lines I_A and I_B whereas the true helical paths are continuously curved. Thus, the correction employed to P_{approx} in (1) is required. This correction was derived in the following manner: (1) use the P_{approx} to calculate the pitch of a perfect coiled coil with pitch, P_1, and axial rise per seven resides, D; (2) make a second order approximation (in D/P_1) of the resulting expression (a closed-form solution for P_1 in terms of P_{approx} is not possible); (3) replace all D/P_1 terms with D/P_{approx}; (4) rearrange the expression to get P_1 as a function of P_{approx} and D. Expression (A1) is accurate to within 0.01 Å when the pitch is around 150 Å. This error becomes smaller for larger pitch values and does not become larger than 1 Å until the pitch drops below 50 Å. The second expression for pitch (A2) is based on a direct interpretation of pitch as the axial distance required before a residue in the same heptad position (residue n to residue n + multiple of 7) is rotated once around the supercoil axis.

Both methods were adapted and used for coiled coils with more than two chains. For method 1, the angle θ_{AB} and the radial separation were measured for each strand relative to the coiled-coil axis. For method 2, the radial vectors were drawn from the coiled-coil axis to each α-carbon position; the angle θ_{12} was then measured for each strand using their respective radial vectors. For a given heptad, the local pitch was measured for each strand individually and averaged.

For an antiparallel coiled coil, the residues of all chains were renumbered so that n and $n+7$ residues on each chain were in axial register.