Protein Science (1995), *4*:2252–2260. Cambridge University Press. Printed in the USA.

The Alacoil: A very tight, antiparallel coiled-coil of helices*

KIM M. GERNERT,[1] MARK C. SURLES,[2] THOMAS H. LABEAN,[1] JANE S. RICHARDSON,[1] AND DAVID C. RICHARDSON[1]

[1] Department of Biochemistry, Duke University, Durham, North Carolina 27710
[2] Interactive Simulations, Inc., San Diego, California 92121

(RECEIVED May 30, 1995; ACCEPTED August 14, 1995)

Abstract

The Alacoil is an antiparallel (rather than the usual parallel) coiled-coil of α-helices with Ala or another small residue in every seventh position, allowing a very close spacing of the helices (7.5–8.5 Å between local helix axes), often over four or five helical turns. It occurs in two distinct types that differ by which position of the heptad repeat is occupied by Ala and by whether the closest points on the backbone of the two helices are aligned or are offset by half a turn. The aligned, or ROP, type has Ala in position "d" of the heptad repeat, which occupies the "tip-to-tip" side of the helix contact where the Cα–Cβ bonds point toward each other. The more common offset, or ferritin, type of Alacoil has Ala in position "a" of the heptad repeat (where the Cα–Cβ bonds lie back-to-back, on the "knuckle-touch" side of the helix contact), and the backbones of the two helices are offset vertically by half a turn. In both forms, successive layers of contact have the Ala first on one and then on the other helix.

The Alacoil structure has much in common with the coiled-coils of fibrous proteins or leucine zippers: both are α-helical coiled-coils, with a critical amino acid repeated every seven residues (the Leu or the Ala) and a secondary contact position in between. However, Leu zippers are between aligned, parallel helices (often identical, in dimers), whereas Alacoils are between antiparallel helices, usually offset, and much closer together. The Alacoil, then, could be considered as an "Ala anti-zipper." Leu zippers have a classic "knobs-into-holes" packing of the Leu side chain into a diamond of four residues on the opposite helix; for Alacoils, the helices are so close together that the Ala methyl group must choose one side of the diamond and pack inside a triangle of residues on the other helix.

We have used the ferritin-type Alacoil as the basis for the de novo design of a 66-residue, coiled helix hairpin called "Alacoilin." Its sequence is: cmSPDQWDKE AAQYDAHAQE FEKKSHRNng TPEADQYRHM ASQY QAMAQK LKAIANQLKK Gsetcr (with "a" heptad positions underlined and nonhelical parts in lowercase), which we will produce and test for both stability and uniqueness of structure.

Keywords: Alacoil; coiled-coil; helix contacts; protein design; side-chain packing; zipper

The work reported here has two distinct origins: one in the generalized study of helix-packing interactions, and the other in the current challenges of de novo protein design. Both sets of background considerations are briefly summarized here.

The formation of contacts between α-helices is an important part of the assembly of protein tertiary structure. Several general descriptions have been given of how side chains on the surface of helices can fit together to achieve particular geometries of helix contact (Crick, 1953; Richmond & Richards, 1978; Efimov, 1979; Chothia, 1984). Analysis in terms of "knobs-into-holes," "ridges-into-grooves," or other packing descriptions generally include three major categories of helix–helix geometry distinguished by the dihedral angle Ω between the helix axes: near +20°, near −60°, and near-perpendicular. High-angle contacts often come as close as 8 Å if there is a Gly or Ala at the central position of the contact on both sides, but the helix axes are about 10 Å or more apart in low-angle contacts. In the coiled-coil arrangement seen extensively in fibrous proteins, the low-angle contact is repeated at each pair of layers, and the two parallel helices curve gently around one another to maintain the same distance and geometry. The helices in globular proteins are usually shorter and straighter, so they diverge somewhat away from their region of closest contact. Low-angle helix contacts in glob-

Reprint requests to: Jane S. Richardson, Department of Biochemistry, Duke University, 211 Nanaline Duke Building, Duke Medical School, Box 3711, Durham, NC 27710; e-mail: jsr@suna.biochem.duke.edu.

*Part of this work was presented by Jane and David Richardson at their AMGEN Award lecture at the First European Symposium of the Protein Society in Davos, Switzerland, on May 31, 1995.

ular proteins are usually antiparallel, although the leucine zipper (Landschulz et al., 1988) turned out to have coiled, parallel helices (O'Shea et al., 1991) just like the fibrous coiled-coils.

The de novo design of protein structures has demonstrated that it is possible to rationally design sequences that will fold to approximately the intended three-dimensional structure at both the secondary and tertiary structure levels (e.g., DeGrado et al., 1989; Hecht et al., 1990; Fedorov et al., 1992; Quinn et al., 1994, etc.) and even to do so with extremely simple hydrophobic-patterning criteria (Hodges et al., 1981; Kamtekar et al., 1993). Designs with all-leucine interiors (e.g., α4 or the coiled-coil mimics) can even be much more stable than natural proteins, but none of the full-length designs show the degree of unique, well-ordered structure typical of natural, natively folded proteins (Richardson et al., 1992; Betz et al., 1993). Rather, they appear to be "molten" by some or all of the following criteria: poor protection from NH exchange, low cooperativity of unfolding, binding of ANS dye, poor chemical-shift dispersion of NMR resonances, a lack of long-range NOEs, and an inability to crystallize. Some long-range order may be present in designs that are either very small (Hill et al., 1990) or are built around large, rigid non-protein groups (Robertson et al., 1994). Especially noteworthy are two helix hairpins whose NMR spectra show half a dozen long-range NOEs between the two helices (Fezoui et al., 1994; Kuroda et al., 1994), evidence that part of the contact attains a specific packing. This overall state of affairs for protein design is loosely analogous to the actual process of protein folding, in which it is rather common to see rapid collapse to a fairly compact, approximately correct structure with a much slower annealing process to a fully ordered native structure (e.g., Ptitsyn et al., 1990). Although there is presently controversy over the correct strategy for success in the second stage of protein design, it must include some components of both "negative design" (building in sequence features primarily aimed at blocking alternative structures) and improvement of internal packing geometry.

Therefore, for our work on the design of helix-bundle proteins, in addition to an emphasis on "negative design" to help ensure the placement, connections, ends, and topology of helices (Hecht et al., 1990; Richardson et al., 1992), we have also undertaken a study of the side-chain and side-chain–backbone packing in proteins, by breaking the structure down into small interlocking units, called "puzzle pieces." We studied those puzzle-piece pairs that occur preferentially at the narrowest parts of low-angle helix contacts (Gernert, 1994). To our surprise, it turned out that the most common single arrangement found for helix-axis separations <8 Å can actually occur as an antiparallel coiled-coil repeat for at least four or five turns of tight contact. This structure, called an Alacoil, is described here.

Results

A survey was made of the geometrical parameters and amino acid identities found for pairs of contacting helical turns (or puzzle-piece pairs) in the low-angle, antiparallel helix–helix contacts of known protein structures (see the Methods). Of those with a helix-axis separation of 8.5 Å or less, about 20% were anomalous because the close contact was at or beyond the end of at least one of the two helices. For the remaining internal, tight helix contacts, nearly all fell into one of two types (described later), each requiring an Ala or other very small residue at a particular position in the contact. These examples are listed in Table 1, along with several distance and angle parameters.

We found no clear cases that could be described as having just one interacting turn-pair with this geometry, so that the shortest examples described here have two adjacent layers of contact. Their helix axes usually make an angle of about +15°, with the required Ala on one helix in one of the layers and on the other helix in the second layer; there is an approximate local twofold relationship between the two layers. From this region of tight contact the helices diverge somewhat, so that other contact layers have wider separations and larger side chains. An alternative arrangement is also found occasionally: with two contact layers, helix axes at about −35°, and the critical Ala or small residues both on the same helix (e.g., 2CTS S261–A265; 3TS1 A299–T303).

Further, about half of the cases listed in Table 1 occur in runs of four or five adjacent contact layers, making up tight, coiled-coil helix contacts, or "Alacoils." One can imagine taking a helix pair with only two tightly contacting layers (as above), putting an Ala in each critical i, $i+7$ position, and then letting the helices coil slightly around each other so that they make tight contacts at each layer. Presumably the favorable contacts at each layer of the coil would compensate for bending the helices.

Geometry of Alacoils

To understand the geometry of these tight helix contact "puzzle pieces," remember that the Cα–Cβ bonds on an α-helix do not point out radially from the helix axis but form a pinwheel, as illustrated in Figure 1A. Viewed from the helix N-terminus, the pinwheel arms point clockwise, leading forward along the N- to C-terminal chain direction. For a pair of antiparallel helices viewed end-on, one will pinwheel clockwise and the other counterclockwise, so that the Cβ spokes mesh like gearwheels (see Fig. 1A). As first emphasized by Efimov (1979), this means that the two sides of an antiparallel helix–helix contact are different (although for the parallel case, both sides are equivalent). On the incoming or "tip-to-tip" side of an antiparallel contact (top, in Fig. 1A, where the chain direction is entering the contact) the Cα–Cβ bonds point approximately toward each other, like the fingers touching at their tips in Figure 1B. On the outgoing or "knuckle-touch" side (bottom of Fig. 1A, where the chain is leaving the contact) the Cα–Cβ bonds point outward and back-to-back, like fingers touching along the backs of their knuckles, as in Figure 1B.

An α-helix normally has a pitch of 3.6 residues per turn, but in a coiled-coil, two helices wind shallowly around each other, producing an effective pitch of 3.5 residues per turn and an exact repeat of the contact geometry every 7 residues, echoed by an approximate 7-residue repeat in the sequence. Each position in this 7-residue "heptad repeat" (McLachlan & Stewart, 1975) has a distinct geometry relative to the other helix; they are named by the letters "a" through "g" for parallel coiled-coils, which terminology we will extend also for use with the present antiparallel cases. One way to describe that distinct geometry is with the "heptad angle," defined as follows: given a residue, construct a plane through its Cα perpendicular to its local helix axis; in that plane, measure the angle Cα–axis–second helix axis. Position "d" has a heptad angle near −40°, whereas position "a" is near +15°. Positions "a" and "d" are the critical hydrophobic ones in the helix contact; they lie on adjacent turns. In an Ala-

Table 1. *Examples of Alacoils*[a]

Protein	"a"	"d"	Δaxis (Å)	Ω (°)	Cβ–Cβ	Offset
Ferritin-type Alacoil ("a" site dominant)						
1FHA ferritin	A 66	E 27	8.1	15	4.0	0.23
	S 31	E 62	7.8	12	3.8	0.26
	S 59	Y 34	7.6	17	3.8	0.25
	S 38	F 55	7.6	19	4.2	nd
	A 52	F 41	7.6	15	4.1	0.24
1VSG variant surface glycoprotein	A 76	M 42	7.7	20	3.8	0.26
	A 46	Y 72	7.4	24	3.6	0.25
	A 69	A 49	7.6	25	3.4	0.24
	A 53	A 65	7.7	21	4.2	nd
	A 62	N 56	7.8	22	4.6	0.19
1LIGb Asp receptor	T 56	H 103	7.9	20	3.7	0.21
	A 100	L 59	7.9	19	3.6	0.22
	T 63	T 96	8.1	17	4.7	0.23
	A 93	N 66	8.4	14	4.2	0.24
	S 70	L 89	8.9	14	4.2	nd
1PRCm photosynthetic reaction center	A 116	I 68	9.2	20	4.7	0.23
	A 65	F 119	7.7	22	4.2	0.22
	S 123	F 61	7.6	21	3.6	0.28
	A 58	S 126	7.6	19	3.8	0.28
2CCY cytochrome *C′*	A 41	A 100	8.2	3	3.9	0.22
	A 97	A 44	7.7	10	3.8	nd
	A 48	S 93	7.7	13	4.0	0.20
	A 90	A 51	8.3	9	3.9	0.20
Adjacent ferritin-type puzzle-pairs, Ω = +15°						
256B cytochrome *b*562	A 87	L 78	7.3	8	4.1	0.23
	A 75	A 90	8.6	5	4.3	0.27
3LZM T4 lysozyme	A 129	M 120	7.9	13	3.8	nd
	S 117	N 132	9.0	12	4.8	0.31
1COLa colicin	S 16	I 45	7.9	−5	4.4	0.20
	A 42	I 19	8.1	−1	3.9	nd
1PRCl photosynthetic reaction center	A 96	F 41	8.1	25	4.1	nd
	A 38	S 99	7.9	20	3.8	0.26
2CTS citrate synthase	A 263	I 145	8.1	10	4.2	0.23
	S 142	M 266	9.0	5	4.3	0.21
ROP-type Alacoils ("d" site dominant)						
1ROP Rop protein	L 9	C 52	9.5 Å	17	5.1 Å	0.45
	Y 49	A 12	8.9	20	4.3	0.44
	R 16	A 45	8.5	18	4.5	0.40
	H 42	T 19	8.6	15	4.1	0.40
	L 23	C 38	8.7	16	4.6	0.34
	A 35	L 26	8.3	15	4.2	0.32
3HHR human growth hormone	E 174	A 17	9.3	20	4.5	0.42
	H 21	M 170	8.4	23	4.0	0.48
	R 167	A 24	8.4	29	4.4	0.46
Adjacent ROP-type puzzle-pairs						
1VSG variant surface glycoprotein	A 94	A 249	8.5	31	4.6	0.46
	A 246	T 97	8.5	31	4.6	0.45
1LPE apolipoprotein A3	T 67	A 29	8.2	15	4.3	0.35
	W 26	E 70	8.0	12	3.9	0.36

[a] For each layer of the Alacoil, residues in the "a" and "d" heptad positions on the two helices are listed (note that they switch from one helix to the other in adjacent layers). Δaxis is the separation between the helix axes, measured locally, and Ω is the dihedral angle between them. Cα–Cβ is the distance between the carbons of residues "a" and "d." Offset is the vertical offset (measured as a fraction of the heptad repeat) between "a" residues on the two helices; if "nd," it was too close to a helix end for measurement.

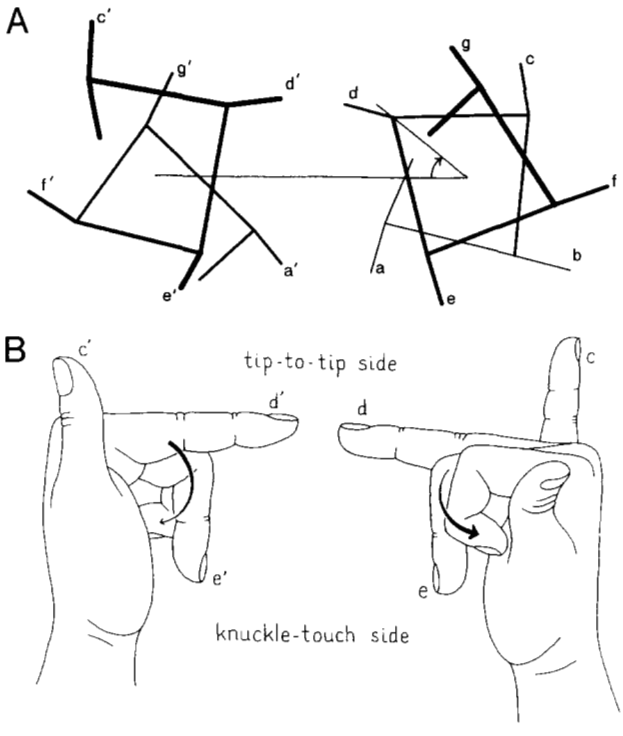

Fig. 1. A: Simplified view down a coiled-coil of antiparallel α-helices. Cα–Cβ vectors and Cα–Cβ bonds are shown, with the positions of the heptad repeat labeled "a"–"g" for the helix at right. Note that the bonds pinwheel clockwise on the lefthand helix and counterclockwise on the righthand one, so that those bonds point toward one another on the top face of the helix contact (the incoming, or tip-to-tip, side) and extend back-to-back on the bottom face (the outgoing, or knuckle-touch, side). The definition of the "heptad angle" is indicated for position "d"; in this case, the indicated angle is negative, because it comes before the helix midline in the N-to-C sequence direction. **B:** Schematic drawing of how two hands can mimic the side-chain contact geometry of a pair of antiparallel helices, with the tip-to-tip and knuckle-touch sides of the contact labeled. Helix backbone direction is shown by perspective arrows, and overall orientation matches A. (Note that the left hand uses the fingers in reverse order, in order to stand in for a right-handed helix. Note also that positions "d" and "d'" are not at the same depth.)

coil, each contact layer is equivalent, with an "a" residue touching a "d," whereas the classic coiled-coil packs the residues symmetrically, but in alternating layers, "a" with "a" and "d" with "d" (see comparison in Table 2). For the antiparallel case, position "a" is on the knuckle-touch side of the contact and position "d" on the tip-to-tip side. These differences in interaction geometry, as well as the closer helix spacing, explain why Ala is preferred in Alacoils but is rather unfavorable in classic coiled-coils (Hodges et al., 1981). In Table 1, the two residues listed for each puzzle-piece pair are heptad position "a" on one helix and heptad position "d" on the other helix; they switch sides in alternate layers.

The vertical offset of the contacting helices is measured by choosing the Cα of the "a" heptad position on one helix as the reference zero level and position "a + 7" as level 1.0; the relative height of the "a" position on the other helix gives the offset (as a fraction of the seven-residue, two-turn repeat). For parallel coiled-coils (including leucine zippers), the axis separation is about 9.6 Å, the vertical offset is 0, and equivalent heptad positions lie directly opposite each other on the two helices. For the present antiparallel Alacoil examples, the closest axis separation is 7.5–8.4 Å, and the offsets found are either 0.4–0.5 or else close to 0.25. An offset of 0.5 produces helix turns that lie directly across from each other, so that they look aligned, but those turns come from different halves of the heptad repeat. An offset of 0.25 or 0.75 gives helix backbones that look evenly interdigitated when viewed from the side (they are offset by half a turn of the two-turn repeat), but the 0.75 case was not found.

Packing of the Ala side chain

The Alacoil is a variant of the classic "knobs into holes" packing of more standard coiled-coils, in which each side chain on one helix fits into a diamond of four side chains ($i-3$, i, $i+1$, $i+4$) on the surface of the other helix (Crick, 1953). For the Alacoil, the helices are so close together that the alanine methyl group must choose one triangular side of the diamond in which it will be centered. That Ala Cβ touches the backbone and either two or all three side chains of the triangle (see Fig. 3B; Kinemage 3). Two types of Alacoil described below can be thought of as determined by whether the Ala chooses to fit into the first triangle ($i-3$, i, $i+1$; ferritin-type) or the second triangle (i, $i+1$, $i+4$; ROP-type) of the diamond.

The critical Ala position can also be occupied by Ser or Cys, or sometimes even Thr, but any residue with a δ-carbon will not fit (in the ferritin type). At first glance that seems puzzling, because it looks as though there is room for a longer side chain to reach out of the interface. However, modeling the possible conformations for a Cδ (see Kinemage 3), it can be seen that:

Table 2. *Comparison of coiled-coil types*

	Ferritin-type Alacoil	ROP-type Alacoil	Classic coiled-coil[a]
Helix directions	Antiparallel	Antiparallel	Parallel
Helix-axis spacing	7.5 Å	8.5 Å	9.6 Å
Fraction offset/heptad	0.2–0.25 (offset)	0.4–0.5 ("aligned")[b]	0 (aligned)
Main contact site	Ala in "a"	Ala in "d"	Leu in "d"
Contact layer(s)	Layers same: a-d	Layers same: d-a	Layers alternate: d-d, a-a
Contact side(s)	Sides different: tip versus knuckle	Sides different: tip versus knuckle	Sides same
Hole opposite knob	Triangle up: $i-3$, i, $i+1$	Triangle down: i, $i+1$, $i+4$	Diamond: $i-3$, i, $i+1$, $i+4$
Occurrence as long coil	In five unrelated globular or membrane proteins	In two unrelated globular proteins	Leu zip dimers; fibrous proteins

[a] Detailed parameters for the classic coiled-coil are taken from the GCN4 leucine zipper structure
[b] Helix turns are aligned, although two-turn heptad repeat is offset by 0.5.

(1) if $\chi 2$ is near $-60°$ it will hit the Cβ of the "e" heptad residue on the opposite helix (so that would work only if position "e" were a Gly); (2) if $\chi 2$ is near 180°, it hits the backbone O from the preceding turn of the opposite helix (so that would work only if the opposite helix began with the contact layer being considered); and (3) if $\chi 2$ is near $+60°$, it hits the backbone of its own helix. Therefore, a larger side chain means that the two helices will be forced further apart.

Ferritin-type versus ROP-type

For the two types of especially tight antiparallel helix–helix puzzle pieces found in our survey, the "ferritin type" has alanines in heptad position "a," helix–axis spacing of close to 7.5 Å (see Table 1), and a vertical offset of about 0.25; whereas the "ROP type" has alanines in heptad position "d," helix–axis spacing near 8.5 Å, and an offset of about 0.5. Table 2 summarizes the properties of the two types of Alacoils and contrasts them with the classic parallel coiled-coils. Ferritin (Lawson et al., 1991) contains a good example of the first, more common, type of Alacoil, whereas the ROP protein (Banner et al., 1987) monomer is a good example of the second type. Figure 2A and Kinemage 1 show several ferritin-type Alacoils superimposed, whereas Figure 2B and Kinemage 2 show similar views for the ROP-type Alacoils, each with the contacting "a" and "d" side chains included. In the two figures, the "a"–"d" contact direction is emphasized by a white line. The difference between the two Alacoil types is most dramatic from the side, where helix backbone turns of the ferritin type look alternately interdigitated, whereas those of the ROP type lie directly across from one another.

Typically, four side chains of a puzzle-piece pair are close enough to potentially touch: "a" and "g" positions on one helix and "d" and "e" on the other, as seen for the individual superimposed helix–turn pairs in Figure 3A or Kinemage 3. In all these examples, the packing geometry matches extremely closely, leading to very close superpositions. For the ferritin-type Alacoil, 70% of the critical, tightly-packed "a" positions are occupied by Ala, 25% by Ser, and the rest by Thr; Cys is probably possible but has not been observed. In the ROP-type Alacoils, the critical "d" position is somewhat more variable, even though fewer examples have been identified; this seems reasonable given the slightly wider helix–axis separation. It can apparently accom-

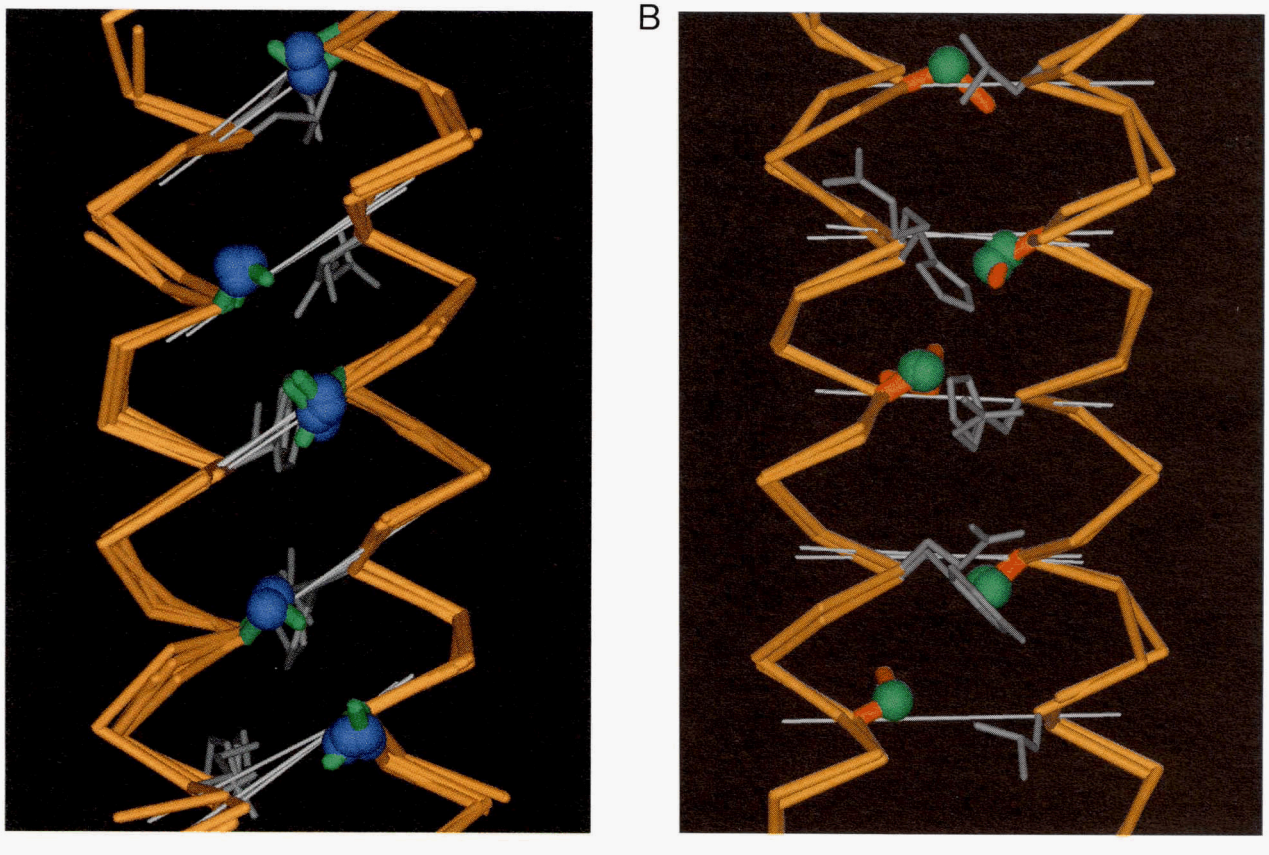

Fig. 2. A: Four long examples of the ferritin-type Alacoil structure, superimposed in side view. The Cα backbones are shown in gold, plus the small side chains (Ala or Ser, in green, with a blue ball at the Cβ position) for the critical "a" heptad position that packs tightest against the opposite helix, and gray side chains for the "d" position. Thin white lines join the "a" Cα to a point halfway between "d" and "e," to show the diagonal nature of the closest side-chain contacts. Note that the helix turns appear to interdigitate: the vertical offset between equivalent "a" positions on the two helices is 0.25 times the heptad repeat, or half of a helix turn. **B:** The only two long examples of the ROP-type Alacoil structure superimposed in side view, showing Cα backbones in gold, the small side chains in the critical "d" heptad position of tightest packing in orange with green balls at the Cβ atoms, and the "a" side chains opposite them in gray. Thin white lines join the "d" Cα to a point midway between "g" and "a." Contacting turns lie directly opposite each other on the two helices, but the vertical offset of heptads is 0.5, so that one "d" position occurs at each turn pair.

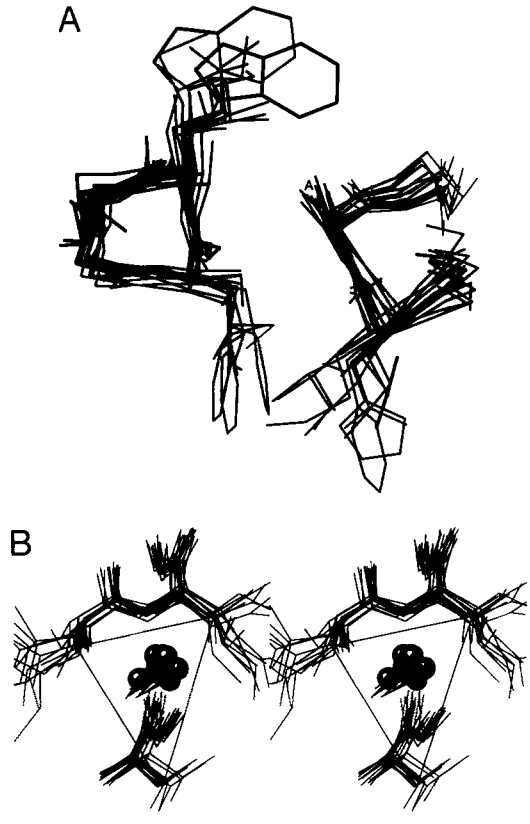

Fig. 3. A: End view of 15 individual helix–turn pairs of the ferritin-type Alacoil superimposed, with side chains shown for the four closest residues. The critical "a" position is labeled "A"; its alanine methyl group packs tightly against the other helix. Ser is allowed in the "a" position, but any side chain with Cα or longer forces the helices significantly further apart. **B:** Stereo of the same 15 ferritin-type Alacoil contacts as in A, but viewed from the Ala side chain (Cβ atoms marked with a ball). The triangle of three residues on the opposite helix is indicated, into which the Ala Cβ nestles.

modate Ala, Thr, Cys, and a scattering of other amino acids. In the overall structure of ROP protein, an especially notable feature is that two such Alacoils form a four-helix-bundle dimer, with a hydrophobic core of four-residue layers made of the two alanines from the anti-zipper interactions plus two leucines or similar residues in the somewhat wider dimer interface (Banner et al., 1987). Recently, a successful redesign of the ROP interior idealized each of those layers to be two Ala and two Leu (Munson et al., 1994).

Environments of Alacoils

Most of the Alacoil helix pairs are buried on one side and exposed to solvent on the other side, but there seems to be no strong preference for whether the tip-to-tip or the knuckle-touch side of the helix contact is the exposed one. Positions "d," "e," and "g" are most often either large hydrophobics (Phe, Leu, Ile, Met) or large polars (Tyr, Arg, Lys, Glu, Gln), with some Ala; a representative selection can be seen in Figure 3A or Kinemage 3. The cytochrome *C′* example has alanines in both the "a" and "d" positions (4 Ala in "a"; 3 Ala and 1 Ser in "d"), so that the local sequence of that helix pair looks consistent with either a ferritin-type or a ROP-type Alacoil. It is actually ferritin-type, but that could be either because of a preference in the local structure or because that geometry fits better with the rest of the four-helix bundle.

Although many of the Alacoils are connected as helix hairpins, those connections are usually rather long and/or include awkward ϕ,ψ values. There may not exist a favorable, neat, short loop that connects two helices with this relative geometry.There are also antiparallel helix–helix contacts that resemble the ones shown here but include larger side chains and have a wider helix separation. Their geometry and patterns of sidechain interaction are much more variable, however, because they are freed from the constraints of fitting the Ala methyl into the triangle of residues on the opposite helix. Also, it is common for Alacoils (including the classic ones in ferritin and ROP) to have additional helix–turn pairs at one or both ends but with wider geometry.

Discussion

The Alacoil has several unusual features that make it of interest for general considerations of protein structure. It puts the two helices significantly closer together than any previously described type of coiled-coil, and yet it is fairly common, with long examples so far identified in seven unrelated proteins. The individual contacts are very tightly fitted together and closely reproducible within each of the two types, with each alanine methyl group touching the backbone of the opposite helix inside a triangle of three residues. In spite of the conformational determinism of the Alacoil, however, it should have lower inherent stability than a leucine-based coil, because at least half of the primary contact residues are Ala (or Ser, Thr, etc.) with relatively low hydrophobicity. However, hydrophobic side chains on an outer face of the Alacoil can take part in other interactions, as happens, for instance, in the extremely stable ROP protein.

Because the achieving of unique, well-ordered structure and the possible tradeoffs between stability and uniqueness are currently central issues in the field of de novo protein design, experimental exploration of the properties of Alacoils could make a useful contribution. We have therefore designed an antiparallel coiled-coil hairpin of 66 residues, called "Alacoilin," based on more than seven layers of the ferritin-type Alacoil and using the Sculpt program for modeling (Surles et al., 1994). The Alacoilin sequence is: cm**S**PDQWDKE A**A**QYDAH**A**QE FEKK**S**HRNng TPE**A**DQYRHM A**S**QYQAM**A**QK LKAI**A**N QLKK **G**setcr, with the "a" heptad positions underlined, the helix N- and C-caps in boldface, and the nonhelical parts in lowercase. The design process is described in the Methods and in Kinemage 4. Figure 4 shows a schematic model, and Figure 5 and Kinemage 5 a detailed model of the expected Alacoilin structure. We will produce and characterize the Alacoilin molecule, especially in terms of stability and uniqueness of structure.

Methods

The analysis of helix contacts, including dividing them into small interlocking packing units called "puzzle pieces," was done using the VIEW exploratory visualization system (Bergman et al., 1993) on a Silicon Graphics 4D/440 workstation. A new set of VIEW scripts was written for this purpose, allowing trial of al-

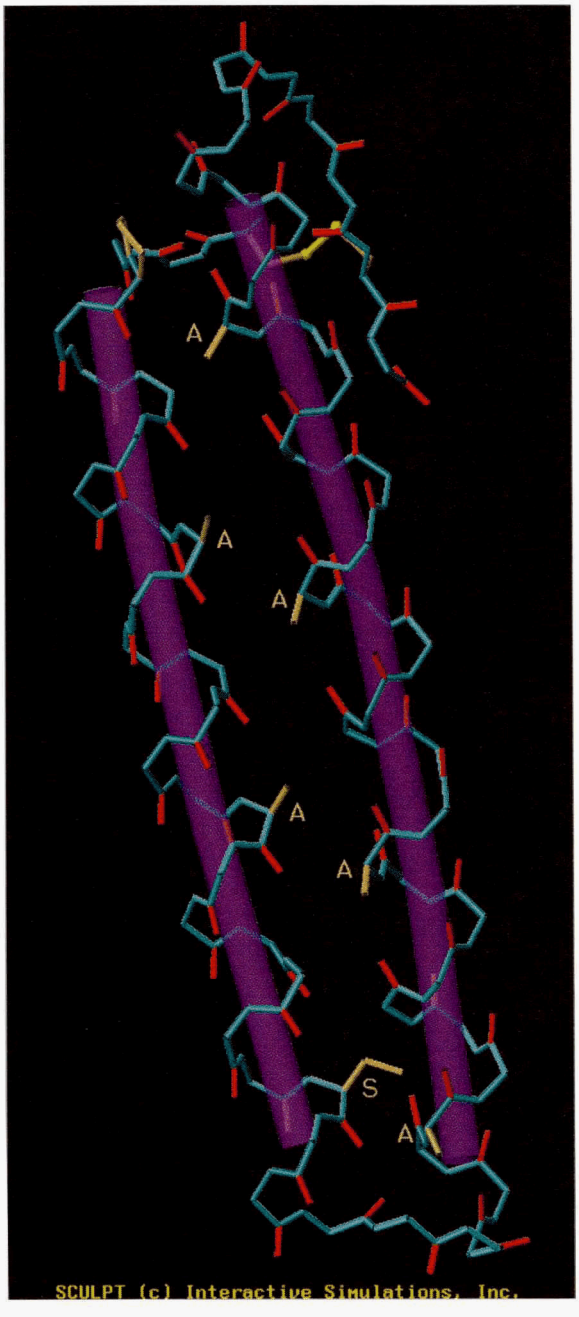

Fig. 4. Simplified view of the Alacoilin designed protein, with purple cylinders along the helices and labels at the critical "a" heptad positions. Alacoilin is based on the ferritin-type Alacoil, with more than seven layers of contact and 66 residues. The single disulfide that joins the two chain termini can be seen at the top rear.

ternative methods for specifying the location of contacting turn pairs and the geometrical parameters for describing them. The final set of definitions used for this part of the study is described in the Results. The primary set of proteins examined was drawn from the July 1993 release of the Brookhaven Protein Data Bank (Bernstein et al., 1977), augmented by some later all-helix structures. Helix pairs at low contact angles were initially located using the Define-S program (Richards & Kundrot, 1988). Puzzle-piece pairs with similar geometry and/or amino acids were examined both in VIEW, superimposed on one puzzle piece and comparing the opposite one, and in Mage running on a Macintosh (Richardson & Richardson 1992, 1994), superimposed on both puzzle pieces. Figure 2A and B was made with VIEW, Figure 4 is taken from Sculpt, and Figures 1A, 3, and 5 and the kinemages were made with Mage. The broader study of helix–helix puzzle pieces is described in Gernert (1994), whereas the current work concentrated on that subset of contacts with helix axes closer than 8.5 Å.

The design of Alacoilin was done using Sculpt (Surles et al., 1994), a program that augments interactive modeling with real-time energy minimization, so that the molecule responds to the user's tugs and movements in a physically realistic way. Kinemage 4 illustrates how the Alacoilin design proceeded, by animating development of the model through a succession of steps.

The highly reproducible geometry of Alacoils makes specification of a backbone framework for the coiled-coil portion of Alacoilin extremely straightforward. The Alacoil example from ferritin was superimposed onto itself with positive and negative offsets of seven residues (first step of Kinemage 4), and those three coordinate sets were combined and idealized to produce a coiled-coil pair of eight-turn helices (nearly four complete heptad units) with almost perfectly repeating geometry. Onto that framework were then built the helix caps, connection, chain termini, disulfide, and side chains, obeying statistical correlations of conformation and local sequence while using Sculpt simultaneously to achieve good packing and favorable dihedral angles.

In the Alacoilin helix–helix contact, almost all of the "a" positions were chosen to be Ala, with one Ser, and one Pro. That Pro provides a highly preferred N1 residue for the first helix (Richardson & Richardson, 1988) and has no preceding helical turn to collide with the Pro ring. The Ser was chosen at the salt-link design stage (see below). Tyr is the most used residue in the "d" position, where it can provide strong hydrophobic contacts close to the helix and a polar group at the outer end; in order to avoid exact repeats in the Alacoilin sequence, some of the "d" positions have other amino acids, such as Trp, Phe, or Leu. Those aromatics, together with surrounding side chains like Gln or Met, form a partially hydrophobic layer on the tip-to-tip side of the Alacoilin helix-pair, but with polar atoms on much of its surface.

Because none of the natural Alacoil hairpin connections are both short and classic, the first attempt at forming a connection in Alacoilin tried to use classic helix "capping boxes" (Harper & Rose, 1993; Seale et al., 1994) taken from those found in other proteins, and then to connect their ends, but Sculpt soon showed that such a connection would stick out sideways from the helices by a surprising and unacceptable amount. Therefore, we decided to use those capping boxes as the two chain termini and connect the central hairpin of Alacoilin with the same conformation that joins the Alacoil in 1VSG. To ensure that Alacoilin would not form a single long helix and then dimerize (see discussion in Hecht et al. [1990]), we checked that the hydrophobic stripes, or the heptad repeats, would be offset by two residues in such a hypothetical long helix. The hairpin connection does not cover any of the coil surface, but at the other end of Alacoilin, the two chain termini were extended far enough to protect some of the coil surface under them and provide a small hydrophobic core centered on the single Trp residue, W7 (see Fig. 5; Kinemage 5). Those termini are joined by a disulfide,

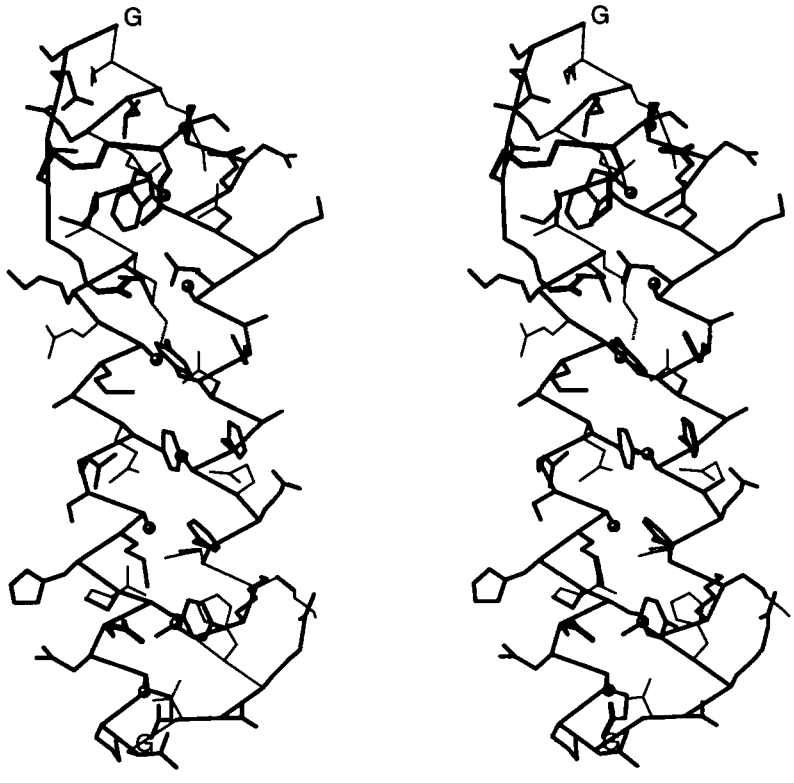

Fig. 5. Cα and side-chain model of the Alacoilin design, viewed from the opposite side (the incoming, "d" residue, or "tip-to-tip," side) to that in Figure 4. Cβ atoms of the critical "a" heptad positions are emphasized with balls. The single Trp side chain can be seen near the top, with the disulfide above and in front of it. The hairpin connection (at bottom) is similar to one in 1VSG, and the chain ends have classic helix capping-box arrangements (Seale et al., 1994) and extend far enough to provide hydrophobic protection for part of the coiled-coil surface. The rest of the surface is designed to be hydrophilic.

C1-C65, whose geometry both matches one of the common SS conformations (Richardson, 1981) and also fits well into the preferred backbone conformations of the chain termini. All four helix ends use amino acids with high single-position occurrence frequencies at the N-cap or C-cap and adjacent three positions (Richardson & Richardson, 1988).

Sequences for the connection and ends were chosen to be as nonhomologous as possible to their origin protein, given the needs both of that local conformation and of the environment within Alacoilin. The N-terminus was changed from LSEGEW (1MBO 2-7) to MSPDQW, the connection from EINHGTNR (1VSG 54-61) to HRNNGTPE, and the C-terminus from IESGKDV (4FXN 24-30) to LKKGSET. There is still some strain remaining in the designed helix-helix connection of Alacoilin, and we also suspect that incompatible requirements between tightly designed connections and packing of the rest of the tertiary structure may contribute to the difficulty of producing well-ordered designed proteins. Therefore, we have sculpted several alternative Alacoilin connections of differing length, including some intended to be quite flexible. Also, if Alacoilin is not sufficiently stable, another heptad can be inserted into each helix without altering the other interactions. If possible, several of these variants will be made and tested.

Possible salt links between the two helices could work best on the knuckle-touch side of the contact (where they can lie across Ala "a" rather than Tyr "d" residues), in heptad position "e." Possible combinations of side-chain identity and conformation for each relative geometry were modeled with Sculpt and were checked for conformer, H-bond geometry, and contact with underlying residues (salt links seen in traditional coiled-coils are not useful models because of the closer helix spacing and different geometry here). The best pair found was an Asp and a Lys in "e" positions across a long diagonal, both with *trans* χ_1 angles, and with Ala for both underlying "a" positions four residues back in sequence from the Asp and the Lys. Also possible is a Glu-Arg pair in "e" positions across a short diagonal, with Ser at the underlying "a" position three residues past the Glu. Two pairs of the first type (D8-K59, D15-K52) and one of the second (E22-R38) were used on the knuckle-touch side of Alacoilin, each with a distinct environment of surrounding residues. Three salt links also join the chain termini to each other and to the coiled-coil (Nterm-E63, E10-R66, K50-Cterm).

The rest of the sequence (mostly heptad positions "b" and "f") was chosen to have a diverse composition of polar amino acids that prefer helix, with some favorable side-chain H-bonds along each helix (either modeled individually or taken from Klingler & Brutlag [1994]), an asymmetrical charge distribution that complements the helix dipole (Shoemaker et al., 1987), and avoidance of repeating sequences. No sequence triplets are repeated, and 19 of the 20 amino acids are used (all but Val). The charge balance was chosen as 10 Lys + Arg, 9 Asp + Glu, and 3 His, giving a calculated pI of 8.25. Secondary structure prediction algorithms strongly predict both helices and the intervening turn (Chou & Fasman, 1978; Garnier et al., 1978).

To check for spurious sequence identity of Alacoilin with natural proteins an MPsrch search (Sturrock & Collins, 1993) of the SwissProt database (release 30) was obtained from the EMBL server (BLITZ@ EMBL-Heidelberg.DE) using the best local similarity algorithm (Smith & Waterman, 1981). Using PAM matrices 60, 90, and 120 (Dayhoff et al., 1978) and default gap penalties, no statistically significant matches were detected. The highest scores resulted from a short six-residue match with a group of viral RNA polymerases (e.g., RRPO_PPMVS and RRPO_TMV) whose structures have not been solved.

Long helices in myosin and HLA class I histocompatibility antigen gave very weak sequence matches of 11–16% over stretches of 20–25 residues. None of the proteins used as partial starting points in the design of Alacoilin (ferritin, myoglobin, flavodoxin, and variant surface glycoprotein) were returned as hits in the searches. In addition, a search of the SwissProt database using the Darwin system at CBRG (crbg@inf.eth.ch) returned no matches at all.

Acknowledgments

This work was done with the support of NIH research grant GM-15000, NASA contract NRA-91-OSSA-18 for K.M.G., and the use of the Macromolecular Graphics Shared Resource of the Duke Comprehensive Cancer Center. We thank Larry Bergman for help with new functions in VIEW.

References

Banner DW, Kokkinidis M, Tsernoglou D. 1987. Structure of the ColE1 Rop protein at 1.7 Å resolution. *J Mol Biol 196*:657–675.

Bergman LD, Richardson JS, Richardson DC, Brooks FP Jr. 1993. VIEW—An exploratory molecular visualization system with user-definable interaction sequences. *Comput Graphics 27*:117–126.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Betz SF, Raleigh DP, DeGrado WF. 1993. De novo protein design: From molten globules to native-like states. *Curr Opin Struct Biol 3*:601–610.

Chothia C. 1984. Principles that determine the structure of proteins. *Annu Rev Biochem 53*:537–572.

Chou PY, Fasman GD. 1978. Empirical prediction of protein conformation. *Annu Rev Biochem 47*:251–276.

Crick FHC. 1953. The packing of α-helices: Simple coiled-coils. *Acta Crystallogr 6*:689–697.

Dayhoff M, Schwartz RM, Orcutt BC. 1978. In: Dayhoff M, ed. *Atlas of protein sequence and structure, vol. 5* (suppl. 3). pp 345–352.

DeGrado WF, Wasserman ZR, Lear JD. 1989. Protein design, a minimalist approach. *Science 243*:622–628.

Efimov AV. 1979. Packing of α-helices in globular proteins. Layer-structure of globin hydrophobic cores. *J Mol Biol 134*:23–40.

Fedorov AN, Dolgikh DA, Chemeris VV, Chernov BK, Finkelstein AV, Schulga AA, Alakhov YuB, Kirpichnikov MP, Ptitsyn OB. 1992. De novo design, synthesis and study of albebetin, a polypeptide with a predetermined three-dimensional structure. Probing the structure at the nanogram level. *J Mol Biol 225*:927–931.

Fezoui Y, Weaver DL, Osterhout JJ. 1994. De novo design and structural characterization of an α-helical hairpin peptide—A model system for the study of protein folding intermediates. *Proc Natl Acad Sci USA 91*: 3675–3679.

Garnier J, Osguthorpe DJ, Robson B. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J Mol Biol 120*:97–120.

Gernert KM. 1994. Structural characterization and redesign of Felix, a designed four-helix bundle protein [thesis]. Durham, North Carolina: Duke University.

Harper ET, Rose GD. 1993. Helix stop signals in proteins and peptides: The capping box. *Biochemistry 32*:7605–7609.

Hecht MH, Richardson JS, Richardson DC, Ogden RC. 1990. De novo design, expression, and characterization of Felix: A four-helix bundle protein of native-like sequence. *Science 249*:884–892.

Hill CP, Anderson DH, Wesson L, DeGrado WF, Eisenberg D. 1990. Crystal structure of α1: Implications for protein design. *Science 249*:543–546.

Hodges RS, Saund AK, Chong PCS, St-Pierre SA, Reid RE. 1981. Synthetic model for two-stranded α-helical coiled-coils. *J Biol Chem 256*:1214–1224.

Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH. 1993. Protein design by binary patterning of polar and nonpolar amino acids. *Science 262*:1680–1685.

Klingler TM, Brutlag DL. 1994. Discovering structural correlations in α-helices. *Protein Sci 3*:1847–1857.

Kuroda Y, Nakai T, Ohkubo T. 1994. Solution structure of a de novo helical protein by 2D-NMR spectroscopy. *J Mol Biol 236*:862–868.

Landschulz WH, Johnson PF, McKnight SL. 1988. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science 240*:1759–1764.

Lawson DM, Artymiuk PJ, Yewdall SJ, Smith JMA, Livingstone JC, Treffry A, Luzzago A, Levi S, Arosio P, Cesarini G, Thomas CD, Shaw WV, Harrison PM. 1991. Solving the structure of human H ferritin by genetically engineering intermolecular contacts. *Nature 349*:541–544.

McLachlan AD, Stewart M. 1975. Tropomyosin coiled-coil interactions: Evidence for an unstaggered structure. *J Mol Biol 98*:293–304.

Munson M, O'Brien R, Sturtevant JM, Regan L. 1994. Redesigning the hydrophobic core of a four-helix bundle protein. *Protein Sci 3*:2015–2022.

O'Shea EK, Klemm JD, Kim PS, Alber T. 1991. X-ray structure of the GCN4 leucine zipper, a two-stranded parallel coiled-coil. *Science 254*:539–544.

Ptitsyn OB, Pain RH, Semisotnov GV, Zerovnik E, Razgulyaev OI. 1990. Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Lett 262*:20–24.

Quinn TP, Tweedy NB, Williams RW, Richardson JS, Richardson DC. 1994. Betadoublet: De novo design, synthesis, and characterization of a novel β sandwich protein. *Proc Natl Acad Sci USA 91*:8747–8751

Richards FM, Kundrot CE. 1988. Identification of structural motifs from protein coordinate data: Secondary structure and first-level supersecondary structure. *Proteins Struct Funct Genet 3*:71–84.

Richardson DC, Richardson JS. 1992. The kinemage: A tool for scientific communication. *Protein Sci 1*:3–9.

Richardson DC, Richardson JS. 1994. Kinemages—Simple macromolecular graphics for interactive teaching and publication. *Trends Biochem Sci 19*:135–138.

Richardson JS. 1981. Anatomy and taxonomy of protein structures. *Adv Protein Chem 34*:167–339.

Richardson JS, Richardson DC. 1988. Amino acid preferences for specific locations at the ends of α-helices. *Science 240*:1648–1652.

Richardson JS, Richardson DC, Tweedy NB, Gernert KM, Quinn TP, Hecht MH, Erickson BW, Yang Y, McClain RD, Donlan ME, Surles MC. 1992. Looking at proteins: Representations, folding, packing, and design. *Biophys J 63*:1186–1209.

Richmond TJ, Richards FM. 1978. Packing of α-helices: Geometrical constraints and contact areas. *J Mol Biol 119*:537–555.

Robertson DE, Farid RS, Moser CC, Urbauer JL, Mulholland SE, Pidikiti R, Lear JD, Wand AJ, DeGrado WF, Dutton PL. 1994. Design and synthesis of multi-haem proteins. *Nature 368*:425–432.

Seale JW, Srinivasan R, Rose GD. 1994. Sequence determinants of the capping box, a stabilizing motif at the N-termini of α-helices. *Protein Sci 3*:1740–1745.

Shoemaker KR, Kim PS, York EJ, Stewart JM, Baldwin RL. 1987. Tests of the helix dipole model for stabilization of α-helices. *Nature 326*:563–567.

Smith TF, Waterman MS. 1981. Identification of common molecular subsequences. *J Mol Biol 147*:195–197.

Sturrock SS, Collins JF. 1993. MPsrch version 1.5. University of Edinburgh: UK Biocomputing Research Unit.

Surles MC, Richardson DC, Richardson JS, Brooks FP Jr. 1994. Sculpting proteins interactively: Real-time energy minimization embedded in a geometrical modeling system. *Protein Sci 3*:198–210.