Research Article

X-ray Structure of the GCN4 Leucine Zipper, a Two-Stranded, Parallel Coiled Coil

ERIN K. O'SHEA, JULI D. KLEMM,* PETER S. KIM, TOM ALBER

The x-ray crystal structure of a peptide corresponding to the leucine zipper of the yeast transcriptional activator GCN4 has been determined at 1.8 angstrom resolution. The peptide forms a parallel, two-stranded coiled coil of α helices packed as in the "knobs-into-holes" model proposed by Crick in 1953. Contacts between the helices include ion pairs and an extensive hydrophobic interface that contains a distinctive hydrogen bond. The conserved leucines, like the residues in the alternate hydrophobic repeat, make side-to-side interactions (as in a handshake) in every other layer of the dimer interface. The crystal structure of the GCN4 leucine zipper suggests a key role for the leucine repeat, but also shows how other features of the coiled coil contribute to dimer formation.

TRANSCRIPTION FACTORS OF THE RECENTLY IDENTIFIED bZIP class, such as C/EBP, Fos, Jun, CREB, and GCN4, regulate the expression of many different genes in organisms as diverse as fungi, plants, and mammals (*1*). The activities of bZIP proteins are determined both by the recognition of specific DNA sequences and by the stability and specificity of protein dimer formation.

The key to dimerization is a conserved sequence motif called the leucine zipper, and contacts with DNA are made by an adjacent region rich in basic residues (*1, 2*). Studies of synthetic peptides have shown that leucine zipper sequences, although only 30 to 40 residues long, are sufficient for dimerization of the GCN4 protein (*3*) and for specific heterodimer formation by Fos and Jun (*4*). Leucine zipper sequences, originally identified on the basis of a heptad repeat of leucines (*1*), are now known to fold as short, parallel coiled coils (*3–6*).

The coiled coil, in which α helices wrap around each other in a shallow left-handed supercoil, was one of the first protein structures modeled on the basis of x-ray fiber diffraction evidence (*7–9*). The coiled coil motif has attracted continued attention because it is found in many proteins, including muscle proteins, α-keratin, bacterial surface proteins, intermediate filaments, laminins, dynein, tumor suppressors, and oncogene products (*10, 11*). Coiled coils also have been identified as ideal candidates for protein design (*10, 12, 13*). In fibrous proteins, coiled coils generally form extended ropes that are several hundred angstroms long. These molecules have proven difficult to crystallize, and a high-resolution x-ray crystal structure of a parallel, two-stranded coiled coil has yet to be obtained.

The general architecture of the parallel coiled coil, however, is well characterized. Crick proposed in 1953 that the dimeric structure could be stabilized by the packing of "knobs" formed by the hydrophobic side chains of one helix into "holes" formed by the spaces between side chains of the neighboring helix (*8*). Consistent with this model, hydrophobic residues are spaced every four and then three residues apart (a 4,3 hydrophobic repeat) in the primary sequences of coiled coils (*10, 14, 15*). This pattern defines a heptad repeat, $(abcdefg)_n$, in which the generally hydrophobic residues at positions a and d fall on the same face of a helix. In parallel coiled coils, oppositely charged residues commonly occur at positions e and g of adjacent heptads, which is consistent with the formation of interhelical ion pairs (*10, 15*). These patterns of hydrophobic and charged residues are also apparent in leucine zipper sequences, with the conserved leucines occurring at position d of the heptad repeat (*3*).

We report the 1.8 Å x-ray crystal structure of a peptide corresponding to the leucine zipper of the transcriptional activator GCN4. GCN4 is responsible for the general control of amino acid biosynthesis in yeast (*16*). Distinct regions of the protein are required for transcriptional activation, DNA binding, and dimerization (*17*). Dimerization, which is required for DNA binding, depends on the leucine zipper sequence in the last 33 residues of the protein. This sequence alone was incorporated into the 33–amino acid peptide (GCN4-p1) described here (Fig. 1).

Structure determination. The initial electron density map of GCN4-p1 was calculated at 3 Å resolution with phases based on the isomorphous and anomalous differences of a single $PtCl_4$ derivative (Table 1). After solvent flattening (*18*), FRODO (*19*) was used to trace 29 amino acids of one helix in the electron density. Because the remaining electron density was discontinuous, the second helix was generated by rotating a polyalanine representation of the initial model around the noncrystallographic twofold symmetry axis (*5, 20*). After rigid body and positional refinement (*21*), an improved electron density map was calculated by using the model phases to resolve the twofold ambiguity of phases based on the isomorphous differences alone (*22*). This electron density map was subjected to eight cycles of twofold averaging, solvent flattening, and phase resolution. The final averaged map was used to build two models of the GCN4-p1 dimer that differed in the register of the sequence in the electron density. The models were refined against 6 to 2 Å data by using simulated annealing methods (*21*), and one of the models yielded a significantly lower crystallographic *R* value and was

E. K. O'Shea is in the Howard Hughes Medical Institute and Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, and in the Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139. J. D. Klemm is in the Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, UT 84132. P. S. Kim is in the Howard Hughes Medical Institute and the Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, and in the Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139. T. Alber is in the Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, UT 84132, and in the Department of Chemistry, University of Utah, Salt Lake City, UT 84112.

*Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

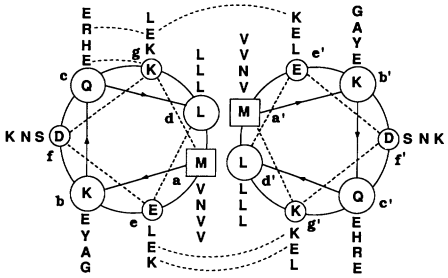

Fig. 1. Helical wheel representation of residues 2 to 31 of GCN4-p1. View is from the NH_2-terminus, and residues in the first two helical turns are boxed (Met^2) or circled. Heptad positions are labeled a through g. Leucines at position d interact across the interface with residues at d′ and e′. Alternate layers contain residues a, a′, g, and g′. Residues that form ion pairs in the x-ray crystal structure are connected with dashed lines. GCN4-p1 corresponds to residues 249 to 281 of the GCN4 protein (*3, 47*). Residues of the adjacent basic region that are needed for DNA binding are not present in the peptide.

chemically reasonable. This model was improved by several rounds of rebuilding and refinement against 6 to 1.8 Å data (*21, 23, 24*).

The current model contains the NH_2-terminal 31 of 33 residues of both polypeptide chains and 52 water molecules. The following data support the correctness of the structure: (i) The R value is 0.179 for 2σ data from 6 to 1.8 Å resolution. The 2σ data represent greater than 84 percent of the reflections in this resolution range. (ii) The overall root-mean-square (rms) deviations from ideal bond lengths, and bond angles are 0.018 Å and 2.5°, respectively. (iii) All backbone dihedral angles are within allowed regions of the Ramachandran plot (*25*), and most side chain dihedral angles are near the preferred rotamers (*26*). (iv) The model fits the $2F_o - F_c$ map well (Fig. 2), and the $F_o - F_c$ map contoured at $\pm 3\sigma$ has no interpretable features. (v) The $F_{PH} - F_P$ difference Fourier calculated with the model phases produces 9 to 10σ peaks at the expected Pt sites, and these sites are within 3 Å of the sulfur atoms of Met^2 in each chain.

Overall fold and dimensions of the molecule. The GCN4 leucine zipper peptide forms a two-stranded, parallel coiled coil of helices (Fig. 3). The dimer is a twisted elliptical cylinder ~45 Å long and ~30 Å wide. The helices wrap around each other to form approximately ¼ turn of a left-handed supercoil (Fig. 3A). The pitch of the supercoil averages 181 Å, and the average distance between the helix axes is 9.3 Å (*27*). This constant separation is maintained by the occurrence of residues of similar size along the length of the interface. The crossing angle of the helices is 18° (Fig. 3B), which matches Crick's prediction for the crossing angle of helices in coiled coils (*8*). The superhelix axis of GCN4-p1 is nearly straight.

The first 30 residues of each peptide monomer form more than eight helical turns. Gly^{31} is not in a helical conformation in either chain of the dimer, and Glu^{32} and Arg^{33} are not visible in the electron density map. The crystallographic B values are generally higher at the helix termini. The average main-chain dihedral angles for residues 3 to 30 of each helix are $-63° \pm 7°$ for ϕ and $-42° \pm 7°$ for ψ. The dihedral angles cluster near the average values of $-63°$ and $-42°$ seen in helices in globular proteins (*28*). There is no apparent correlation of ϕ, ψ values with position in the heptad repeat.

The individual helices are smoothly bent, which permits tight contacts over the length of the dimer. The curvature is associated with shorter main chain hydrogen bonds in the interface compared to the outside of the helices (Fig. 4). In particular, hydrogen bonds from the amides of residues at position e of the heptad repeat tend to be shorter, whereas the amides of residues at position f form longer helical hydrogen bonds (*29*). Pauling and Corey proposed that such a difference in hydrogen bond lengths could cause supercoiling of helices (*9*). As in solvent-exposed helices in globular proteins (*28*), the main chain carbonyl groups of surface residues of GCN4-p1 are often hydrogen-bonded to ordered water molecules.

The helices in the leucine zipper are related by an approximate local twofold rotation axis. The α-carbons of residues 1 to 30 of each monomer can be superimposed with an rms deviation of 0.64 Å (*30*). Overall, conformational differences between the helices are as large as differences between heptads within a given helix (for example, see Fig. 4). Consequently, both the particular sequence of each heptad and the distinct environment of each peptide monomer in the crystal are likely to contribute to the observed local structural variations.

Dimer interface. The packing of side chains in the dimer interface conforms to Crick's knobs-into-holes model (*8*) (Fig. 5). The leucines (at position d) and the amino acids in the alternate hydrophobic position (a) are surrounded by four residues from the neighboring helix. This packing is related by a translation of the helices to the "ridges-into-grooves" scheme that describes most helix-helix contacts in globular proteins (*31*). In ridges-into-grooves packing, however, each residue in the interface makes contact with only two residues of the neighboring helix. In contrast, the pattern of four side chains surrounding each residue at positions a and d of the leucine zipper maximizes buried surface area and likely contributes to the considerable stability of the dimer.

The leucine zipper dimer also can be represented as a twisted ladder in which the sides are formed by the helix backbones and the rungs are formed by side chains in the interface (Fig. 3C). The conserved leucines are not interdigitated, but instead they make side-to-side interactions in every other rung. In alternate rungs, side-to-side contacts are made by residues in position a of the heptad repeat.

The layers of the interface, however, contain four residues, not two (Fig. 5). Each conserved leucine at position d packs against both the symmetry-related leucine (d′) and the side chain of the following residue (e′) (Fig. 6A). In adjacent layers (Fig. 6B), the amino acid at position a packs between its symmetry mate (a′) and the preceding residue (g′). These two types of layers alternate through the structure and form an extensive hydrophobic interface

Table 1. Data collection and phasing statistics. Crystals (space group $C2$) of the GCN4-p1 peptide, Ac-RMKQ LEDKVEE LLSKNYH LENEVAR LKKLVGE R [conserved leucines bold and alternate hydrophobic residues underlined (*45*); Ac, acetyl], were obtained from 25 mM phosphate, 0.4 M NaCl (pH 7.2) by using polyethylene glycol as the precipitant (*5*). X-ray data were collected [with a Xuong-Hamlin area detector at the University of California at San Diego (*46*)] from one crystal each of the native peptide and a K_2PtCl_4.

Parameter		Native	K_2PtCl_4
Unit cell dimensions (Å)	a	101.6	100.4
	b	30.4	30.3
	c	21.9	21.9
(degrees)	β	94.7	94.5
Measured reflections		11,527	22,494
Unique reflections		5,790	10,878
R_{merge}*		0.043	0.038
R_{iso}†			0.20
Completeness of data to 1.8 Å resolution (percent)		92	87
Number of sites			2
Rms F_H/E‡ (20 to 3Å)			2.04
R_{FHLE}§			0.49
R_{Cullis} ‖			0.73
Mean figure of merit (20 to 3 Å)			0.73

*$R_{merge} = \Sigma|I - \langle I\rangle|/\Sigma I$; I, intensity. †$R_{iso} = \Sigma|F_{PH} - F_P|/\Sigma F_P$; F_{PH} and F_P, derivative and native structure factor amplitudes, respectively. ‡Rms $F_H/E = [(\Sigma f_H^2)/\Sigma(F_{PH}(obs) - F_{PH}(calc))^2]^{1/2}$; F_H = heavy-atom structure factor amplitude, and f_H = heavy-atom scattering factor. §$R_{FHLE} = \Sigma|F_{HLE} - F_H(calc)|/\Sigma F_{HLE}$. ‖$R_{Cullis} = \Sigma|F_H(obs) - F_H(calc)|/\Sigma F_H(obs)$.

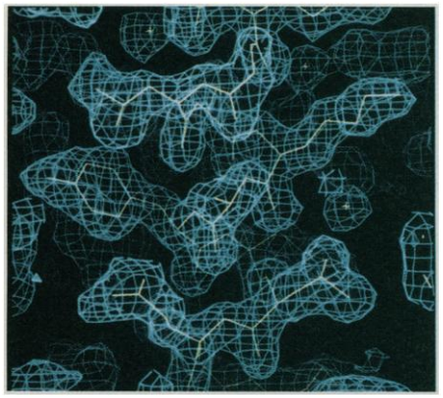

Fig. 2. A portion of the 6 to 1.8 Å resolution $2F_o - F_c$ electron density map superimposed on the current model. A side view of residues 12 to 20 is shown.

(Fig. 6C). Approximately 1800 Å^2 of surface area is buried upon forming the dimer from helical monomers; >95 percent of this surface area is from the side chains of residues at positions a, d, e, and g (*32*). The side chains of residues at positions a and d are 83 percent buried in the dimer.

All valines at position a and all leucines at position d adopt the most preferred rotamer conformations [$\chi_1 \sim -60°$, $\chi_2 \sim 180°$ for Leu, and $\chi_1 \sim 180°$ for Val (*26*)], which aligns the branched Leu side chains along the superhelix axis and facilitates interhelical contacts between layers in the interface. The γ-methyl groups of Val^9, for example, interact with Cδ2 of Leu$^{5'}$ (4.4 Å) and Cδ1 of Leu$^{12'}$ (3.8 Å) in the neighboring helix. These contacts with adjacent layers are not equivalent. The alternate hydrophobic residues are generally closer to the succeeding leucines than to the preceding leucines, as might be expected because the 4, 3 repeat contains two different spacings.

The buried positions a and d are structurally distinct: side chains at these positions have different orientations relative to the dimer axis (Fig. 6, A and B, and 7). The observed structural differences correlate with distinct sequence preferences of positions a and d (*33–36*).

A distinctive hydrogen bond in the dimer interface appears to be formed between Asn^{16} side chains at position a of the heptad repeat (Fig. 6D). The amide and carbonyl groups of the two Asn^{16} side chains are 2.6 Å apart, and the side chains are in different conformations. This asymmetric model, which apparently is trapped by the crystal lattice, is favored for at least three reasons: (i) The asymmetric structure fits the electron density calculated with phases obtained from a refined model that lacks the side chains of residues 15, 16, and 20 of both helices and three nearby water molecules. (ii) The asymmetry extends to neighboring residues, including Lys^{15} and Glu^{20}. One of the Lys^{15} residues makes an intermolecular contact through a water molecule. (iii) When the Asn^{16} residues are placed in a single preferred rotamer conformation, the model does not fit the electron density and the side chains cannot make favorable contacts in the dimer interface (*37*).

Leucine zippers that lack polar residues at position a or d are rare (*1, 34, 35*), which suggests that buried polar groups like Asn^{16} in GCN4-p1 have important functions. Examination of the crystal structure suggests that Asn^{16} may actually be destabilizing. The Asn side chains bury polar substituents, pack more loosely against adjacent layers than do Val or Met at position a, and, as discussed below, appear to disrupt an interhelical ion pair. Moreover, Asn^{16} is especially tolerant of amino acid substitutions in a chimeric protein in which the leucine zipper of GCN4 mediates dimerization of the DNA binding domain of λ repressor (*38*). Destabilization of the leucine zipper could help make dimerization reversible in vivo, modulate the affinity of bZIP proteins for DNA by controlling the concentration of dimers, or have both effects.

In addition, Asn^{16} may help position the helices in a parallel, unstaggered orientation. Antiparallel or staggered arrangements would be destabilized because the Asn side chains would pack against nonpolar residues. Similarly, polar amino acids in the leucine zipper interface could contribute to the specificity of heterodimer formation by favoring associations of sequences with complementary polar groups.

Electrostatic interactions. Electrostatic complementarity is seen in the structure of GCN4-p1. The net charge of the leucine zipper at neutral pH is near zero (+1), and positive and negative residues generally alternate along the helices. This arrangement permits both intra- and interhelical ion pairing (Fig. 1).

Distances between charged side chains suggest that interhelical ion pairs are formed between Lys^{15} and Glu$^{20'}$, Glu^{22} and Lys$^{27'}$, and Glu$^{22'}$ and Lys^{27}. These pairs of residues occur at position g of one heptad and position e′ of the following heptad in the neighboring helix. Lys^{15} and Glu^{22} (at position g) both precede an alternate hydrophobic residue, and Glu^{20} and Lys^{27} (at position e) follow a conserved leucine. As a result, the methylene groups of residues involved in interhelical ion pairs also help form the hydrophobic core of the dimer (Fig. 6A).

The dual roles of charged residues at positions e and g suggest that electrostatic and hydrophobic interactions in the leucine zipper are interdependent. This idea is consistent with the results of genetic

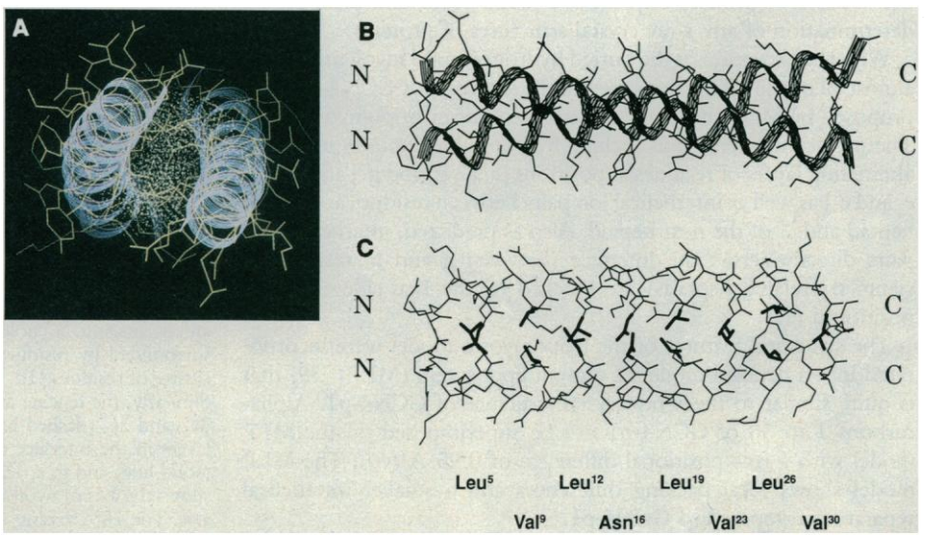

Fig. 3. Views of the GCN4-p1 dimer that illustrate features of the parallel coiled coil. (**A**) View along the superhelix axis from the NH_2-terminus. The main chain is highlighted with a ribbon and the reduced van der Waals surfaces of the side chains at positions a and d are stippled in yellow. The helices are curved, and the overall superhelical twist is ~90°. (**B**) Side view with a ribbon representation of the main chain. The crossing angle of the helices is ~18°. (**C**) Side view perpendicular to (B). The side chains of the residues that make up the 4,3 hydrophobic repeat (positions d and a) are shown in bold. The GCN4-p1 dimer resembles a ladder in which the sides are formed by the backbones of the helices and the rungs are formed by hydrophobic side chains.

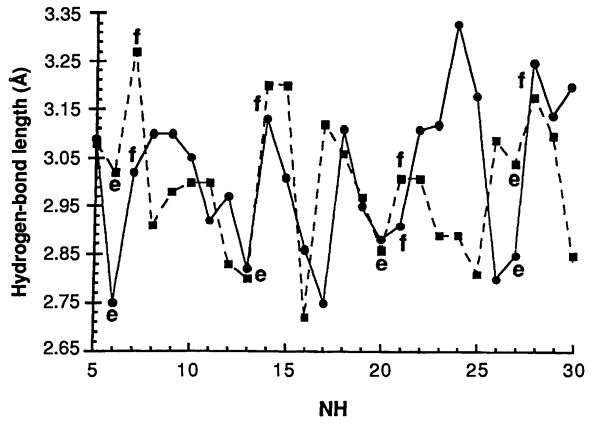

Fig. 4. Helical hydrogen bond lengths in the GCN4-p1 structure. The distance from the main chain amide nitrogen of residue i to the carbonyl oxygen of residue $i + 4$ is plotted as a function of the sequence position of the amide group. The solid and dashed lines represent the two different helices. Residues at position e tend to form shorter hydrogen bonds, whereas hydrogen bonds to amides at position f tend to be longer (*29*).

studies (*38*) that show that Leu^{19} and Leu^{26}, which are bracketed by ion pairs, are less tolerant of amino acid substitutions than Leu^{5} and Leu^{12}. More direct evidence for the influence of packing on electrostatic interactions is found in the GCN4-p1 structure, where $Asn^{16'}$ sterically blocks the formation of an ion pair between Lys^{15} and $Glu^{20'}$. This region of the structure graphically illustrates how the formation of interhelical ion pairs depends on a complementary surface provided by buried residues at positions a and d.

Intrahelical ion pairs are also apparent in the structure of GCN4-p1. In one helix, for example, there are close contacts between Lys^{8} and Glu^{11} (3.3 Å) as well as between Glu^{22} and Arg^{25} (2.8 Å). Glu^{22} is also near $Lys^{27'}$ (3.6 Å) from the adjacent helix, suggesting competition between inter- and intrahelical ion pairs. Fewer intrahelical ion pairs are seen in the crystal structure than anticipated from the sequence. Many of the charged residues (at positions b, c, and f) that are expected to participate in intrahelical ion pairs are involved in crystal contacts.

Comparison to other two-stranded coiled coil structures and models. The high-resolution x-ray crystal structure of the GCN4 leucine zipper confirms earlier models of two-stranded, parallel coiled coils (*8, 15, 39*). As predicted, the helices are crossed at ~18°, packed symmetrically, and stabilized by knobs-into-holes interactions between hydrophobic residues in the dimer interface. It is remarkable that Crick proposed this structure almost 40 years ago in the absence of primary sequences of coiled coils and prior to the determination of any x-ray crystal structures of proteins.

With the exception of the buried hydrogen bond involving Asn^{16}, almost all of the interactions seen in the structure of GCN4-p1 were proposed by McLachlan and Stewart to occur in tropomyosin (*15*). The predicted interactions include hydrophobic contacts involving alternating layers of residues at positions [a, a′, g, and g′] and [d, d′, e, and e′] as well as interhelical ion pairs between residues at g of one heptad and e′ of the next heptad. Also as predicted, interhelical ion pairs directly across the interface (between g and e′ residues in symmetry-related heptads) are blocked by the Leu side chains at position d (*15*).

The structural features of the tropomyosin model were incorporated into a detailed model of murein lipoprotein (MLP) (*39*) that is quite similar to the x-ray crystal structure of GCN4-p1. Alpha-carbons 1 to 30 of GCN4-p1 can be superimposed on the MLP model with a rms positional difference of 0.95 Å (*40*). The MLP model shows local packing differences and a smaller interhelical separation compared to GCN4-p1.

The GCN4-p1 dimer, however, is quite different from the antiparallel coiled coil protein ROP, even though the individual helices of the two proteins superimpose well (*41, 42*). The ROP dimer forms a four-helix bundle; each monomer consists of a pair of supercoiled, antiparallel helices. Identical helices on the corners of the four-helix bundle are parallel, but these helices are more than 4 Å farther apart than the helices in GCN4-p1.

GCN4-p1 nonetheless shows qualitative similarities to antiparallel coiled coils. In the antiparallel coiled coil domain of the *Escherichia coli* seryl-tRNA synthetase, for example, a 4,3 hydrophobic repeat occurs in the sequence, the residues at positions a and d are buried in the interface, and ion pairs are formed within and between the helices (*43*).

Parallel and antiparallel coiled coils are distinguished by distinct sequence patterns that reflect different pairwise interactions in the dimers. In antiparallel coiled coils, residues at a and d′ are paired, as are residues at d and a′. This staggering of the heptad repeats is required to keep the hydrophobic residues in register. In the GCN4 leucine zipper, however, a and a′ residues as well as the leucines at d and d′ are side-by-side in the dimer interface.

Another important difference between parallel and antiparallel coiled coils is the distribution of charged residues. In antiparallel coiled coils, residues at e and e′ occur on one face of the dimer and residues at g and g′ occur on the other face. Complementary charges occur at pairs of e residues and pairs of g residues that are structurally adjacent (*43*). In contrast, the sequences of parallel coiled coils are characterized by oppositely charged residues at positions g and e of the following heptad (*10, 15, 36*). These residues form the interhelical ion pairs in GCN4-p1. The observation of interhelical ion pairs in coiled coil structures suggests that the distinctive distributions of charged residues at positions e and g influence the orientation of helices in the dimer.

Comparison to classical coiled coils. In at least two respects, the GCN4 leucine zipper is an atypical coiled coil. First, the leucine zipper is much shorter than most coiled coils in fibrous proteins. In addition, leucine occurs almost invariably at position d in leucine zipper sequences, whereas in classical coiled coils only one-fourth to one-half of the residues at position d are leucine (*36*).

The differences in length between leucine zippers and other coiled

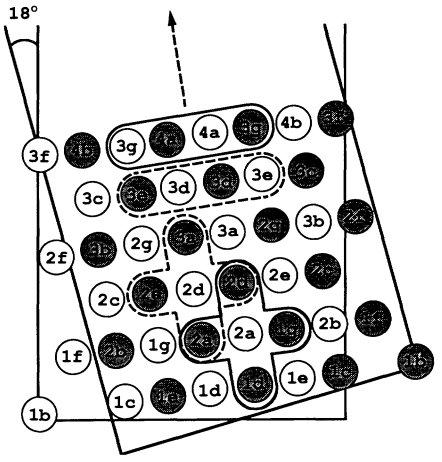

Fig. 5. Helical net diagram of knobs-into-holes packing [adapted from (*8*)]. The diagram is obtained by wrapping a piece of paper around each helix, marking the positions of the α-carbons with circles, and placing one piece of paper on top of the other in a manner reflecting packing of the dimer interface. The Cα atoms of the two helices are represented by open and shaded circles. Heptad positions are indicated a to g. Residue 2a, for example, fits into a "hole" surrounded by residues 1d′, 1g′, 2a′, and 2d′ (solid line). An analogous cluster of residues [1d, 1g, 2a, and 2d (not marked)] surrounds residue 2a′. Similarly, the leucine at 2d forms a "knob" that is surrounded by 2a′, 2d′, 3a′, and 2e′ (dashed line). In the upper part of the figure, examples of the layers in the interface containing the residues at positions e, d, d′, and e′ (solid line) and g, a, a′, and g′ (dashed line) are marked. The dashed arrow shows the axis of twofold rotational symmetry coincident with the superhelix axis. The 18° crossing angle of the helices is indicated at the upper left.

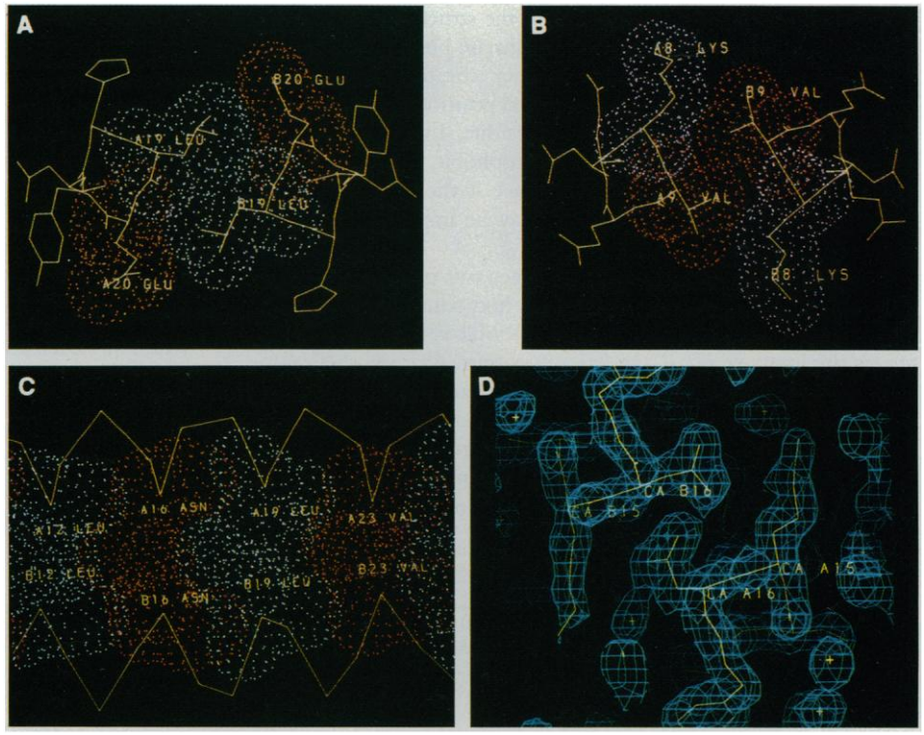

Fig. 6. Sections through the GCN4-p1 structure illustrating interactions that form the dimer interface. (**A**) The conserved leucines make side-to-side contacts and also interact with the succeeding residues at positions e and e′. The van der Waals surfaces of the Leu^{19} residues (positions d and d′) are shown in light blue and the surfaces of the Glu^{20} residues (positions e and e′) are shown in red. Glu^{A20} forms an ion pair with Lys^{B15} (not shown). Equivalent residues from each helix are distinguished by "A" or "B" preceding the residue number. The view is analogous to the helical wheel diagram in Fig. 1. (**B**) In alternate rungs, side-to-side contacts are made by residues of the alternate hydrophobic repeat. The alternate hydrophobic residues (at a and a′) also interact with the preceding residues at positions g and g′. The van der Waals surfaces of the Val^{9} residues (at a and a′) are red and the surfaces of the Lys^{8} residues (g and g′) are purple. View is from the NH_2-terminus, as in (A). (**C**) The interface is continuous and well packed. Side view of the GCN4-p1 dimer showing the van der Waals surfaces of residues at positions a (red) and d (light blue) superimposed on the helix backbones. (**D**) A portion of the $2F_o - F_c$ electron density map showing the region of asymmetry around Asn^{16}. View is from the NH_2-terminus along the axis of the superhelix. The two Asn^{16} side chains are modeled in different conformations (χ_1 values = +68° and −175°). The side chain of Asn^{A16}, which is turned into the interface, is 2.8 Å from a water molecule. The neighboring Lys^{15} side chains are also asymmetrical; Lys^{B15} forms an ion pair with Glu^{A20} (not shown), and Lys^{A15} approaches a water molecule (3.7 Å) that makes an intermolecular contact in the crystal. The charged termini of Lys^{A15} and Glu^{B20} are 6.1 Å apart.

coils are likely to reflect their diverse functions. Traditional coiled coils have dynamic roles in motility and cell structure that often require large surfaces for interactions with multiple proteins over large distances (*10*). In contrast, the leucine zipper motif is primarily a dimerization interface (*1*). Studies of peptide models suggest that four heptad repeats of a coiled coil sequence are sufficient for dimerization (*3, 13, 44*).

Why are leucines conserved at position d? Trivially, the apparent conservation may be a consequence of the use of the heptad repeat of leucines as the primary criterion for identifying bZIP proteins. Leucine zipper dimers may be part of a larger class of proteins that associate through a coiled coil motif. The leucine zipper sequences of cpc-1, TGA1a, and TGA1b, for example, each contain two residues at position d that are not leucine (*34, 35*).

The conservation of leucines, however, suggests that the repeat serves important functions. A much discussed idea is that the leucine repeat is a common adaptor that mediates heterodimer formation. Heterodimers can confer multiple regulatory activities on individual protein chains and may facilitate interactions between different regulatory circuits. The conservation of the leucines implies that variations at other positions determine the relative affinities of different leucine zipper sequences.

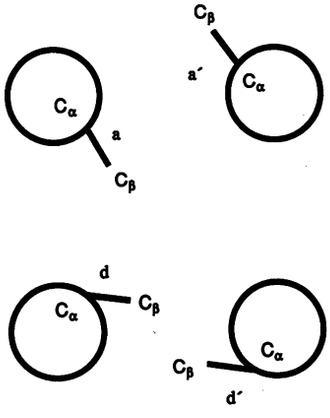

Fig. 7. Schematic drawing (not to scale) showing differences between positions a and d in the GCN4-p1 structure. The Cα-Cβ vectors for positions a and d point in different directions relative to the dimer axis. The conserved leucines are pointed into the interface, and the β-carbons of symmetry-related leucines are almost 2 Å closer together than the β-carbons of equivalent alternate hydrophobic residues. The circles represent the helices in cross section.

The leucine repeat is also almost certainly especially stabilizing. Genetic analysis of the GCN4 leucine zipper shows that the conserved leucines generally are less tolerant of amino acid substitutions than the alternate hydrophobic residues at position a (*38*). In addition, peptide dimers corresponding to a tropomyosin consensus sequence with leucines at positions a and d are destabilized when pairs of leucines are replaced by other hydrophobic residues (*12*).

An explanation for the stabilizing contributions of branched residues in the interface (Leu at position d and the β-branched residues that often occur at a) is provided by the crystal structure of GCN4-p1. Compared to linear aliphatic side chains, the branched residues fill more space between the helices, pack well with adjacent (e and g) residues, and make closer contacts with adjacent layers in the interface. In homodimers, smaller residues (such as Ala) or larger residues (such as Phe, Tyr, and Trp) could produce packing defects in the interface.

Implications for structure and specificity. The GCN4-p1 structure reveals a striking richness of interactions that determine the stability and specificity of protein pairing. The hydrophobic and ionic contacts appear to explain both the requirement for branched hydrophobic residues at positions a and d and the preponderance of long, charged side chains at the adjacent positions e and g. The diversity of interactions reinforces the point that a heptad repeat of leucines, by itself, is not sufficient to mediate dimerization.

Finally, we note that the structure of GCN4-p1 can be used to model interactions in heterologous leucine zippers and in other parallel, two-stranded coiled coils. With different amino acids at each sequence position, the structural details of different coiled coils are likely to vary.

Of special interest is the preferential formation of heterodimers mediated by the leucine zippers of the nuclear oncogene products

Fos and Jun (*2, 4, 34*). On the basis of the structure of GCN4-p1, both packing and electrostatic interactions are predicted to differ in the Fos-Jun heterodimer leucine zipper compared to the two homodimers. Jun has an Asn at the buried position 16 of the leucine zipper, which is paired with a polar residue, Lys, in the Fos-Jun heterodimer. At the next alternate hydrophobic position, Ala^{23} in Jun would leave a cavity in the interface if the local structure is identical to GCN4-p1. The packing may be improved by pairing with Ile^{23} in the heterodimer.

Moreover, repulsive interactions between two pairs of like charges at positions e and g′ in the Fos and Jun homodimers could disrupt the close packing of charged and hydrophobic residues that characterizes the interface of GCN4-p1. In contrast, four interhelical ion pairs could form between e and g′ residues of adjacent heptads in the Fos-Jun heterodimer. Both in terms of destabilizing the homodimers and stabilizing the Fos-Jun heterodimer, packing and electrostatic interactions analogous to those in the GCN4-p1 structure provide attractive mechanisms for determining specificity.

REFERENCES AND NOTES

1. W. H. Landschulz, P. F. Johnson, S. L. McKnight, *Science* **245**, 1759 (1988); C. R. Vinson, P. B. Sigler, S. L. McKnight, *ibid.* **246**, 911 (1989).
2. P. F. Johnson and S. L. McKnight, *Annu. Rev. Biochem.* **58**, 799 (1988).
3. E. K. O'Shea, R. Rutkowski, P. S. Kim, *Science* **243**, 538 (1989).
4. E. K. O'Shea, R. Rutkowski, W. F. Stafford III, P. S. Kim, *ibid.* **245**, 646 (1989).
5. R. Rasmussen, D. Benvegnu, E. K. O'Shea, P. S. Kim, T. Alber, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 561 (1991).
6. T. G. Oas, L. P. McIntosh, E. K. O'Shea, F. W. Dahlquist, P. S. Kim, *Biochemistry* **29**, 2891 (1990).
7. F. H. C. Crick, *Nature* **170**, 882 (1952).
8. ———, *Acta Crystallogr.* **6**, 689 (1953).
9. L. Pauling and R. B. Corey, *Nature* **171**, 59 (1953).
10. C. Cohen and D. A. D. Parry, *Trends Biochem. Sci.* **11**, 245 (1986); *Proteins* **7**, 1 (1990).
11. V. A. Fischetti, *Sci. Am.* **264**, 58 (June 1991); D. Martin-Zanca, S. H. Hughes, M. Barbacid, *Nature* **319**, 743 (1986); H. Bourne, *ibid.* **351**, 188 (1991); J. Groden *et al.*, *Cell* **66**, 589 (1991); K. W. Kinzler *et al.*, *Science* **253**, 661 (1991).
12. R. S. Hodges, N. E. Zhou, C. M. Kay, P. D. Semchuk, *Peptide Res.* **3**, 123 (1990).
13. K. T. O'Neil and W. F. DeGrado, *Science* **250**, 646 (1990).
14. R. S. Hodges, J. Sodek, L. B. Smillie, L. Jurasek, *Cold Spring Harbor Symp. Quant. Biol.* **37**, 299 (1972); A. D. McLachlan, M. Stewart, L. B. Smillie, *J. Mol. Biol.* **52**, 281 (1975); J. A. Talbot and R. S. Hodges, *Acc. Chem. Res.* **15**, 224 (1982).
15. A. D. McLachlan and M. Stewart, *J. Mol. Biol.* **98**, 293 (1975).
16. E. W. Jones and G. R. Fink, in *The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression*, J. N. Strathern, E. W. Jones, J. R. Broach, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982), pp. 181–299.
17. K. Struhl, *Trends Biochem. Sci.* **14**, 137 (1989).
18. B. C. Wang, *Methods Enzymol.* **115**, 90 (1985).
19. T. A. Jones, *J. Appl. Crystallogr.* **11**, 268 (1978).
20. P. M. D. Fitzgerald, *ibid.* **21**, 273 (1988).
21. A. T. Brünger, J. Kuriyan, M. Karplus *Science* **235**, 458 (1987).
22. W. A. Hendrickson, *Acta Crystallogr.* **B27**, 1474 (1971). J. L. Smith and W. A. Hendrickson, in *Computational Crystallography* 209 (1982). W. E. Royer, W. A. Hendrickson, E. Chiancone, *J. Biol. Chem.* **264**, 21052 (1989).
23. D. E. Tronrud, L. F. TenEyck, B. W. Matthews, *Acta Crystallog.* **A43**, 489 (1987).
24. The refinement was started by rebuilding the model into $2F_o - F_c$ and $F_o - F_c$ maps and adjusting atomic positions by simulated annealing with X-PLOR (*21*). The resulting model, with a single isotropic thermal parameter and no solvent molecules, had an R value of 0.268 against data from 6 to 2 Å resolution. The rms deviations from ideal bond lengths and angles were 0.022 Å and 4°, respectively. Subsequent cycles of simulated annealing positional refinement and restrained B refinement gave an R factor of 0.21 against 3σ data from 6 to 1.8 Å resolution. Water molecules were added, and the model was rebuilt into $2F_o - F_c$ and $F_o - F_c$ maps. Refinement was continued with TNT (*23*) to yield a model with excellent geometry (see text) and a crystallographic R value of 0.179 for 6 to 1.8 Å data. The coordinates of the refined model have been deposited in the Brookhaven Protein Data Bank (identification code 2ZTA).
25. C. Ramakrishnan and G. N. Ramachandran, *Biophys. J.* **5**, 909 (1965).
26. J. W. Ponder and F. M. Richards, *J. Mol. Biol.* **193**, 775 (1987).
27. Superhelical parameters were obtained by using a program that fits the Cα backbone of the GCN4-p1 model (residues 3 to 30) with an ideal two-stranded, parallel coiled-coil Cα backbone (P. Harbury and P. S. Kim, unpublished results). The Cα's for an ideal coiled coil were generated by using Crick's parameterization [F. H. C. Crick, *Acta Crystallogr.* **6**, 685 (1953)] and fit to GCN4-p1 by least-squares minimization. The parameters that were varied include the path length of the superhelix (distance traveled along the axis of the monomer helix in one superhelical turn), the radius of the superhelix, and the radius of the monomer helix. The ideal coiled coil Cα backbone superimposes on the GCN4-p1 backbone with an rms deviation of 0.25 Å. The pathlength (s) is related to the radius of the super helix (r) and the angle that the axis of the monomer helices make with the superhelix axis (ϕ) by the expression: $s\sin\phi = 2\pi r$. The interhelical distance is $2r$. The crossing angle is 2ϕ. The superhelical pitch (P) = $s\cos\phi$. Superhelical pitch is the distance traveled along the superhelix axis in one superhelical turn.
28. T. Blundell, D. Barlow, N. S. Borkakoti, J. Thornton, *Nature* **306**, 281 (1983); P. Chakarabarti, M. Bernard, D. C. Rees, *Biopolymers* **25**, 1087 (1986).
29. The nuclear magnetic resonance (NMR) spectrum of GCN4-p1 in solution (*6*) also shows this pattern. Relative to an unstructured peptide, the chemical shifts of amide protons of interior residues (a, d, and e) tend to be shifted downfield (as expected for shorter hydrogen bonds) and the chemical shifts of amide protons at exterior positions (b, c, and f) are shifted upfield (as expected for longer hydrogen bonds. [I. D. Kuntz, P. A. Kosen, E. C. Craig, *J. Am. Chem. Soc.* **113**, 1406 (1991); E. M. Goodman and P. S. Kim, in preparation].
30. The COMPARE program, written by W. A. Hendrickson [*Acta Crystallogr.* **B35**, 158 (1979)] and modified by G. Petsko, was used to superimpose the models and calculate the rms deviations.
31. C. Chothia, M. Levitt, D. Richardson, *J. Mol. Biol.* **145**, 215 (1981).
32. Surface area calculations were performed with the program CHARMM [B. R. Brooks *et al.*, *J. Comput. Chem.* **4**, 187 (1983)] with a probe radius of 1.4 Å.
33. In leucine zipper sequences (*1, 34, 35*), leucines predominate at position d but occur rarely at position a. This contrasts with longer coiled coils, in which leucines occupy positions a and d with ~26 to 47 percent frequency (*36*). The alternate hydrophobic position in leucine zipper sequences is commonly occupied by the β-branched residues Val, Ile, and Thr. Similarly, Ile and Val are found almost three times more frequently at position a than position d in the coiled coil domain of myosin [A. D. McLachlan and J. Karn, *J. Mol. Biol.* **164**, 605 (1983)].
34. T. K. Kerppola and T. Curran, *Curr. Opin. Struct. Biol.* **1**, 71 (1991).
35. M. Gaire, B. Chatton, C. Kedinger, *Nucleic Acids Res.* **18**, 3467 (1990); M. Nishizawa, K. Kataoka, N. Goto, K. T. Fujiwara, S. Kawai, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 771 (1989); T. Hai, F. Liu, W. J. Coukos, M. R. Green, *Genes Dev.* **3**, 2083 (1989); M. J. Guiltinan, W. R. Marcotte, Jr., R. S. Quatrano, *Science* **250**, 267 (1990).
36. D. A. D. Parry, *Biosci. Rep.* **2**, 1017 (1982); J. F. Conway and D. A. D. Parry, *Int. J. Biol. Macromol.* **12**, 328 (1990).
37. NMR studies suggest that the entire GCN4 leucine zipper, including Asn^{16}, is symmetrical in solution (*6*). The NMR and x-ray results can be reconciled if rapid motions in solution produce similar average environments for equivalent residues in the dimer. Alternatively, the Asn^{16} residues could pack in a symmetrical but torsionally strained conformation in solution. In either case, the asymmetry in side-chain and backbone orientations seen in the crystal structure suggest that a particular conformation that is accessible to the molecules in solution is trapped by intermolecular contacts in the crystals.
38. J. C. Hu, E. K. O'Shea, P. S. Kim, R. T. Sauer, *Science* **250**, 1400 (1990).
39. A. D. McLachlan, *J. Mol. Biol.* **122**, 493 (1978).
40. Residues A6 to A36 and B6 to B36 of the MLP sequence were used for the superposition. For comparison, the α-carbons of residues 5 to 26 in the GCN4-p1 monomers can be superimposed on the MLP monomers with rms deviations of 0.56 and 0.63 Å. The monomers of GCN4-p1 superimpose on an ideal straight helix with rms deviations of 0.61 and 0.67 Å.
41. D. W. Banner, M. Kokkinidis, D. Tsernoglou, *J. Mol. Biol.* **195**, 657 (1987).
42. The rms deviations of the positions of equivalent α-carbons in individual helices in GCN4-p1 and ROP range from 0.55 to 0.64 Å.
43. S. Cusack, C. Berthet-Colominas, M. Hartlein, N. Nassar, R. Leberman, *Nature* **347**, 249 (1990).
44. S. Y. M. Lau, A. K. Taneja, R. S. Hodges, *J. Biol. Chem.* **259**, 13253 (1984).
45. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
46. N.-H. Xuong, D. Sullivan, C. Nielsen, R. Hamlin, *Acta Crystallogr.* **B41**, 267 (1985); A. J. Howard, C. Nielsen, N.-H. Xuong, *Methods Enzymol.* **114**, 452 (1985).
47. A. G. Hinnebusch, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6442 (1984).

48. We thank G. Petsko for advice and encouragement through all aspects of this work, A. Brünger for help with X-PLOR, W. Royer for help with symmetry averaging, X. Nguyen-Huu and K. I. Varughese for help with data collection and processing, and W. A. Hendrickson for suggesting the phase resolution technique. We also thank R. Rutkowski for peptide synthesis, D. Tsernoglou for providing the coordinates of ROP, and B. Tidor and P. Harbury for help with calculations of surface area and superhelical parameters. We are grateful to L. McIntosh and B. Tidor for critically reading the manuscript. We are indebted to the MIT Supercomputing Facility, C. Epstein, D. Irish, and the Evans & Sutherland Computer Corporation for access to computers and to David Goldenberg for help producing Fig. 3. E.K.O. is a Howard Hughes Medical Institute Predoctoral Fellow. Supported by grants to P.S.K. from the National Institutes of Health (GM 44162), the Rita Allen Foundation and the Pew Memorial Trust and to T.A. from the American Cancer Society (MV382) and the Pew Memorial Trust.

8 July 1991; accepted 20 September 1991