

# The Leucine Zipper: A Hypothetical Structure Common to a New Class of DNA Binding Proteins

WILLIAM H. LANDSCHULZ, PETER F. JOHNSON, STEVEN L. MCKNIGHT

**A 30-amino-acid segment of C/EBP, a newly discovered enhancer binding protein, shares notable sequence similarity with a segment of the cellular Myc transforming protein. Display of these respective amino acid sequences on an idealized α helix revealed a periodic repetition of leucine residues at every seventh position over a distance covering eight helical turns. The periodic array of at least four leucines was also noted in the sequences of the Fos and Jun transforming proteins, as well as that of the yeast gene regulatory protein, GCN4. The polypeptide segments containing these periodic arrays of leucine residues are proposed to exist in an α-helical conformation, and the leucine side chains extending from one α helix interdigitate with those displayed from a similar α helix of a second polypeptide, facilitating dimerization. This hypothetical structure is referred to as the "leucine zipper," and it may represent a characteristic property of a new category of DNA binding proteins.**

THE MOLECULAR BASIS OF SPECIFIC INTERACTION BETWEEN proteins and DNA has been the subject of intensive study. Proteins that regulate the various functions and metabolism of DNA can be fixed at an appropriate site of action by binding avidly to certain DNA sequences. Proteins that exhibit this property, termed sequence-specific DNA binding proteins, play integral roles in DNA replication, recombination, strand scission, and transcription. Moreover, additional roles for sequence-specific DNA binding proteins will probably be discovered in processes ranging from chromosome segregation during cytokinesis, to the ordered folding and unfolding of chromatin during successive transitions between interphase and mitosis.

The sequence-specific DNA binding proteins that have been studied most extensively are the gene activator and repressor proteins of bacteria. The active forms of these proteins are rotationally symmetric dimers, and their DNA binding sites are rotationally symmetric palindromes. X-ray crystallographic studies have revealed a structural motif common to the DNA binding domains of many bacterial repressors and activators (*1*). This DNA binding motif, termed "helix-turn-helix," is characterized by two successive α helices juxtaposed at approximately 90° by a turn of four amino acids (*2*). Dimerization of this class of bacterial gene regulatory proteins arranges one α helix of each monomer with its analog so that they fit into successive major grooves of DNA.

A second DNA binding motif, common to many eukaryotic gene regulatory proteins, has recently been discovered. The founding member of this second family is a metalloprotein, termed TFIIIA, which acts as a positive regulator of 5*S* RNA gene expression (*3*). The amino acid sequence of TFIIIA has a repeating motif consisting of two closely spaced cysteines followed by two histidines (*4*). Biophysical studies indicate that a zinc ion is tetrahedrally coordinated by the cysteine-histidine motif, and that the coordination complex imparts both stability and DNA sequence-specificity to the TFIIIA polypeptide (*5*). The amino acid sequence diagnostic of this motif, termed the "zinc finger," is being observed in an ever increasing number of eukaryotic gene regulatory proteins (*6, 7*).

Two fundamental concepts have been established from studies on sequence-specific DNA binding proteins. The first, which satisfies long-standing predictions, holds that binding specificity results from direct atomic interaction between amino acid side chains and base pairs in minimally distorted B DNA. There seems, however, to be no general code that specifies the amino acid surface necessary to interact stably with a given DNA surface. In contrast, a second and less intuitively obvious concept has emerged. That is, a limited number of structural motifs "steer" the appropriate amino acid side chains of a protein into the grooves of double helical DNA where they can interact directly with base pairs. These structural motifs are composed primarily of amino acid residues that do not make direct contact with the DNA. Rather, they form three-dimensional "scaf-

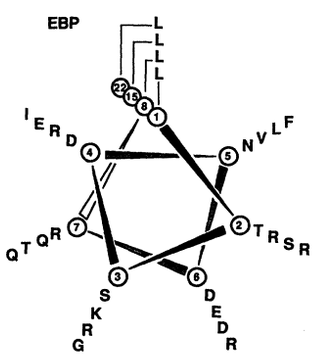



**Fig. 1.** Helical wheel analysis of a carboxyl-terminal portion of C/EBP. The amino acid sequence of a portion of the C/EBP polypeptide (*11*) is displayed end-to-end down the axis of a schematic α helix. The most amino-terminal residue included in the analysis is leucine residue 315 of C/EBP, and is placed at position number one of the idealized helix. The second position on the helical wheel, which is occupied by threonine 316, is located at an azimuth displaced by 100° from position number one. The helical wheel consists of seven "spokes", corresponding to the fit of seven amino acid residues into every two α-helical turns. A total of 28 amino acids of C/EBP are included in the analysis, with the most carboxyl-terminal residue being glutamine 342. The amino acid abbreviations are as follows: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

The authors are in the Department of Embryology, Carnegie Institution of Washington, 115 West University Parkway, Baltimore, MD 21210.

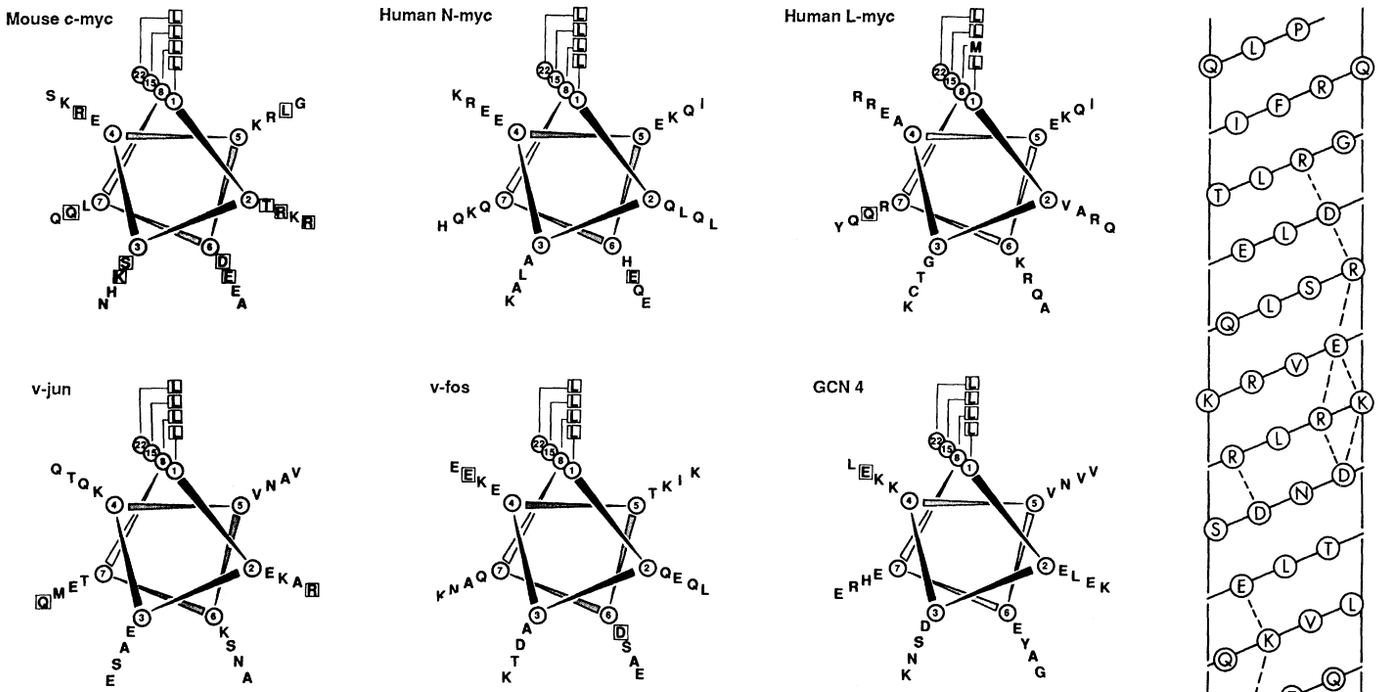



**Fig. 2 (left).** Conservation of leucine repeats between c-Myc, n-Myc, L-Myc, Jun, Fos, and GCN4. The amino acid sequence of a portion of each protein is displayed on a schematic α helix according to the conventions of Fig. 1. The region of c-Myc selected for display had been observed to share substantial sequence similarity with C/EBP (*11*), and starts at leucine residue 413 of the mouse c-Myc polypeptide (*28*). The regions of n-Myc and L-Myc selected for display correspond to the same carboxyl-terminal segment as was chosen for c-Myc. The analysis of n-Myc started at leucine residue 425, and that of L-Myc started at leucine residue 333 (*29*). A slight similarity in amino acid sequence has been observed among c-Myc, Jun, Fos, and the DNA binding domain of GCN4 (*14*). Helical wheel analyses were carried out on these segments starting at leucine residues 465 of Jun (*30*), 165 of Fos (*31*), and 253 of GCN4 (*32*). Amino acid residues that share identity with C/EBP (see Fig. 1) are boxed. **Fig. 3 (right).** Helical net analysis of a carboxyl-terminal region of C/EBP. The helical region containing the leucine repeats of C/EBP was flattened into two dimensions by splitting the helix lengthwise along the face opposite to the strip of leucines. The display starts in the lower left corner with valine residue number 306. The first helical turn is completed with glutamine residue 309, the second turn is initiated by glutamine residue 310, and so on until proline residue 344. The split along the helix bisects two residues, lysine 324 and glutamine 342. As such, each of these residues is displayed in duplicate, corresponding to the "last" and "first" residues of helical turns 5 and 6 and 10 and 11, respectively. Oppositely charged residues located in positions suitable for ion pairing ($i \pm 3$ or $i \pm 4$) are connected by a dashed line.

folds" that match the contour of DNA. These scaffolds dictate the appropriate positioning of the interacting protein surface, allowing atomic interaction between amino acid side chains and the base pairs that constitute a specific binding site on DNA. Indeed, it is the amino acid sequences of these scaffolds, rather than surface-contacting sequences, that exhibit salient protein-to-protein similarity (*6, 8*).

There are ample reasons to anticipate new and different structural motifs for DNA binding proteins. For example, crystallographic studies of the restriction endonuclease Eco RI show that its recognition specificity is not established by either of the aforementioned structural motifs (*9*). Moreover, the amino acid sequences of at least three newly identified sequence-specific DNA binding proteins have failed to show relatedness to either the "helix-turn-helix" motif or the "zinc finger" motif (*10, 11*). Finally, amino acid sequence analyses of nuclear oncogene products, which have been anticipated to modulate gene expression by binding DNA (*12*), have failed to show similarity to either of the established DNA binding motifs.

We now describe an amino acid sequence motif common to five DNA binding proteins: three nuclear transforming proteins, and two transcriptional regulatory proteins. The motif consists of a periodic repetition of leucine residues. We propose that the leucines extend from an unusually long α helix, and that the leucine side chains of one helix interdigitate with those of a matching helix from a second polypeptide to form a stable noncovalent linkage. Furthermore, we predict that the paired helices of this class of proteins play a fundamental role in arranging the contact surface for sequence-specific interaction with DNA.

**A periodicity of leucines.** Rat liver nuclei contain a heat-stable protein that is capable of binding in a sequence-specific manner to regulatory DNA sequences common to a number of different animal virus chromosomes (*13*). Two of the cis-regulatory DNA sequences to which this protein binds are the "CCAAT homology" common to many promoters of genes that encode messenger RNA, and the "enhancer core homology" common to many viral enhancers. Since it was not initially realized that the same protein accounted for both DNA binding activities, it was variously termed CBP (CCAAT binding protein) or EBP (enhancer binding protein). The physiological role of this protein is poorly understood; as such, we continue to designate the protein merely according to its binding specificity (C/EBP).

The gene that encodes C/EBP has been isolated as a recombinant clone, and the DNA binding domain of the protein has been localized to a 14-kD segment (*11*). The amino acid sequence of the C/EBP DNA binding domain contains an abundance of residues with charged side chains, especially lysines and arginines; however, no prolines occur within this region. Since proline residues are seldom found in α helices, we arranged the amino acid sequence of C/EBP on a schematic α helix. When analyzed in this way, a 28-amino-acid segment of the DNA binding domain exhibited notable amphipathy (Fig. 1). One side of the hypothetical helix was

predominately composed of hydrophobic amino acids (particularly leucines), while the other was composed of residues with charged side chains (six basics and four acidics) and uncharged polar side chains (glutamines, threonines, and serines).

Most conspicuous was the periodic repetition of leucine residues. Leucines appear at every seventh position over a region of 35 amino acids within the DNA binding domain of C/EBP. This distribution is not simply a function of the abundance of this residue since the 28-amino-acid region contains only one other leucine. In a computer-assisted search for related protein sequences, we found substantial similarity between C/EBP and the product of the mouse c-*myc* proto-oncogene (*11*). The two proteins share 17 identities within a 30-amino-acid region localized near their respective carboxyl termini. The region of similarity between C/EBP and the transforming protein c-Myc coincides almost perfectly with the proposed α helix of C/EBP shown in Fig. 1. Having noted the unusual repetition of leucines in C/EBP, we arranged the c-Myc sequence on a schematic α helix and discovered the same motif (Fig. 2). The sequence of mouse c-Myc, starting 32 residues upstream from its carboxyl terminus, exhibits leucines at every seventh residue over eight hypothetical turns of α helix. The same periodic array of leucines occurs in human c-Myc, a related human protein termed n-Myc, and at three out of four positions in human L-Myc (Fig. 2). The single deviation from this heptad periodicity of leucines was observed in human L-Myc, and consisted of a replacement by methionine.

Amino acid sequence similarities have been noted between c-Myc and two other nuclear transforming proteins, Fos and Jun (*14*). Moreover, both structural and functional evidence has indicated that Jun is related to GCN4, a DNA binding protein from yeast that plays a direct role in regulating transcription (*14, 15*). When displayed on schematic α helices, we noticed that the related sequences of each of these proteins exhibit at least four periodic repeats of leucine residues (Fig. 2).

The invariant occurrence of at least four leucine repeats in five different proteins is not reflective of general sequence similarity. Although any two of the proteins share substantial similarity within this intriguing region (that is, C/EBP is quite similar to c-Myc, and Jun is even more similar to GCN4), no single amino acid, other than the four leucines, is conserved among all four proteins. Indeed, Jun and GCN4 share only one other identity with C/EBP within this 28-residue window, and Fos shares only two other identities (see Fig. 2).

**Prediction of unusual helix stability**. The leucine repeat common to the aforementioned proteins was observed as a consequence of projecting amino acid sequences on an idealized α helix. If these segments of protein actually exist in an α-helical conformation, the predicted helices would be unusually long. The distance from the first leucine to the fourth, 22 amino acids, would require a minimum of six helical turns. C/EBP, Jun, and Fos actually contain five leucine repeats, which would span at least eight α-helical turns. Two of the primary forces that stabilize α helices are the amphipathic arrangement of hydrophobic amino acids and the occurrence of oppositely charged amino acid pairs configured in a manner that allows formation of a salt bridge (*16–19*). The disposition of hydrophobic residues on one side of an α helix can provide a contiguous array of stabilizing interactions with the globular fold of a protein. In more pronounced cases, hydrophobic interactions of this nature facilitate the coiled-coil intertwining of very long α helices found in keratins, lamins, and paramyosin (*20*).

The stabilizing influence of ion pairs, which depend on the appropriate juxtaposition of acidic and basic amino acids within an α helix (*21*), has been inferred from two lines of evidence. On the analytical side, computer searches of the Brookhaven Data Bank of solved protein structures have shown that α helices are relatively rich in ion pairs. In a computer study of all identified α-helical structures, Sundaralingam *et al.* (*18*) noticed a proportional increase in the frequency of ion pairs as the size of the helix lengthened. The class including the largest α helices, which consisted of at least six helical turns, contained an average of 0.4 ion pair per turn. In a separate study, Sundaralingam also examined the incidence of ion pairs in a category of α helices common to calmodulin and troponin C (*22*). These unusual α helices, which connect dual globular domains within each protein, are exceptionally long, stable, and solvent-exposed; moreover, they are extremely rich in intrahelical ion pairs, containing an average density of 0.7 ion pair per helical turn. On the experimental side, short peptides with systematically varied amino acid sequences have been synthesized and tested for their propensity to adopt solvent-stable α helical structure. Using such an approach, Marqusee and Baldwin (*19*) found that the appropriate juxtaposition of oppositely charged residues fostered helix stability.

As noted above, C/EBP exhibits an amphipathic array of hydrophobic residues in the area that shares sequence similarity with c-Myc (see Figs. 1 and 2). Moreover, this same region of the protein is unusually rich in oppositely charged residues that are juxtaposed in a manner suitable for ion pairing. An α-helical display of the region containing the five leucine repeats of C/EBP (Fig. 3) indicates that eight ion pairs occur within the eight helical turns that separate the first leucine from the fifth leucine. Two additional pairs occur in the helical turns that precede the leucine closest to the amino terminus.

The density of ion pairs observed in C/EBP, roughly one per helical turn and evenly spaced throughout the 30-amino-acid region of interest, is exceptionally high. Only helices known to be long, stable, and solvent-exposed have ion pair densities approaching one per turn. On this basis, we predict that a substantial portion of the DNA binding domain of C/EBP, which includes the leucine repeat motif, exists in a solvent-stable, α-helical conformation. Indeed, since the sequence-specific DNA binding activity of C/EBP is unusually heat-stable (*13*), we suggest that the DNA binding surface may occur within or adjacent to this helix.

**Interhelical interdigitation of leucines: The zipper model**. Why do each of the nuclear proteins that we have considered contain leucine residues at every seventh position over at least six helical turns? The most obvious response to this question is that the leucines establish amphipathy, which helps to stabilize a long α helix. The "helical wheel" plots of Figs. 1 and 2 show that C/EBP, GCN4, and Jun contain hydrophobic residues at three out of four positions on the "spoke" immediately adjacent to that containing the four leucines. This arrangement establishes a continuous "spine" of hydrophobicity over six to eight helical turns. The opposite face of each putative helix is rich in amino acids with either charged side chains or uncharged polar side chains. According to our conventions, "spokes" 3 and 6 are opposite from the "spoke" containing leucine residues. Out of the 40 residues that occupy "spokes" 3 and 6 in the five proteins (C/EBP, c-Myc, Fos, Jun, and GCN4), 21 bear charged side chains and 11 bear uncharged polar side chains.

Although amphipathy is a common feature of each of these putative α helices, the degree of surface hydrophobicity along the nonpolar side is not remarkably high. Two of the proteins have almost no hydrophobic amino acids at "spoke" positions adjacent to the leucines; Fos has only a single isoleucine to abet hydrophobicity, and c-Myc has only a single leucine (both are located at position 3 of "spoke" 5; see Fig. 2). In these cases, surface hydrophobicity is limited to a very thin "ridge" consisting of one leucine residue every other turn of the helix. Furthermore, if hydrophobicity were the only property necessary along the nonpolar side of these helices, it is perplexing that leucines would be used to the virtual exclusion of all other hydrophobic amino acids.

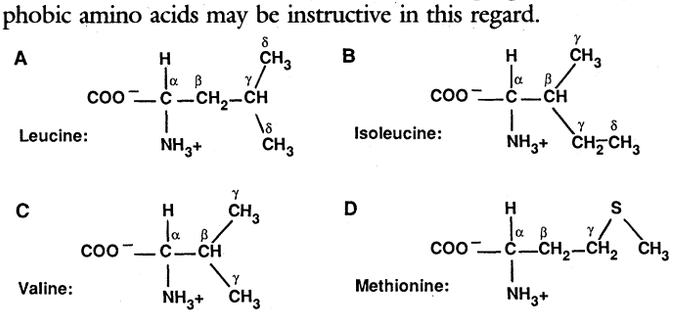
We have used a computer program to generate a molecular model

of C/EBP in the region including its leucine repeats (*23*). The side chains of C/EBP were programmed onto the α carbon backbone of an idealized α helix and displayed in three dimensions. The graphics program highlights van der Waals radii in color. We chose to display the periodically repeating leucine residues of C/EBP in blue and all remaining amino acids in red. Figure 4 shows the putative C/EBP helix as viewed from three different perspectives around the long axis of the helix, each differing by 90°.

Three-dimensional projections of C/EBP show that the van der Waals radii of the repeating leucine side chains are insufficiently close to provide stabilizing intrahelical interactions. This observation reinforced the need to find a hydrophobic surface to match the thin "ridge" of hydrophobicity. Such a surface could exist within the globular fold of the same polypeptide. Alternatively, as in the case of coiled-coil α helices, the complementary surface might be presented by a separate polypeptide. Our attention became focused on the latter possibility for several reasons. (i) The leucine motif maintains strict adherence to a heptad repeat. In no case was the continuity of the repeat shifted by even a single amino acid. The heptad repeat is the quintessential feature shared by all proteins that adopt a coiled-coil structure (*20*). (ii) The hydrophobic surface of the leucine repeat motif seemed unusually thin. If the helix were to exist within a globular fold, the properties of that fold would have to accommodate the unusual aspects of the helix (the narrowness of its hydrophobic "ridge," an abundance of charged amino acids, and an abundance of uncharged polar amino acids).

If we assume that the "ridge" of hydrophobicity of each helix requires a matching surface, and consider that this surface might be donated by a separate polypeptide, perhaps the best candidate is the matching surface donated by the same helix from a second monomer. That is, the hydrophobic surfaces of two leucine helices might interact to form a dimer. Although based on the coiled-coil paradigm, this hypothetical structure is distinctive in several ways. First, the interaction surface is relatively short. Keratins, lamins, and other proteins that adopt a coiled-coil structure rely on dozens or even hundreds of heptad repeats, resulting in extensively interwoven helices (*20*). Moreover, the interhelical hydrophobic interactions of those proteins can be established by almost any hydrophobic amino acid. The motif that we propose covers only four or five heptad repeats, and depends almost exclusively on leucine side chains.

Why is leucine used to the virtual exclusion of all other hydrophobic residues? An examination of the structural properties of hydrophobic amino acids may be instructive in this regard.




The leucine side chain has two methyl groups extending from its single γ carbon, and no methyl groups appended to its β carbon. As such, its side chain is long, symmetrical, and bulky at the tip. We predict that these properties allow the leucine residues from one α helix to interdigitate with those of a second α helix, forming a molecular zipper between two polypeptides (Fig. 5A). It is important to note that an ideal α helix contains slightly more than 3.5 residues per turn. Therefore, if our hypothesis is correct, the paired helices must be distorted in order to maintain maximum packing of the leucine side chains.

Our model of ordered interdigitation may account for the exclusive use of leucine. Like the tooth of a zipper, the leucine side chain is relatively long and bulky at its tip. Isoleucine also contains an extended side chain that is appended with two methyl groups. However, one of the methyl groups is attached to the β carbon of the isoleucine side chain. We propose that the projection of a methyl group from a β carbon would interfere sterically with the sequential interdigitation of hydrophobic side chains. According to these interpretations, valine would constitute a particularly ill-suited amino acid. Its side chain is short (relative to leucine) and is appended with two methyl groups extending from the β carbon atom. Rather than acting to lock two helices together, valine would block interdigitation. Given these considerations, methionine might be expected to constitute the most suitable alternative to leucine. Like leucine, its β carbon is free of attached methyl groups. Moreover, it contains both a methyl group and a bulky sulfur atom at the tip of its side chain. Out of 31 heptad repeats analyzed in our study, we observed 30 leucines and one methionine (see Fig. 2).

The model that we have presented makes the explicit prediction that the "leucine zipper" would represent the dimerization domain of this class of DNA binding proteins. Although this prediction has not been tested directly, it is known that GCN4 and C/EBP exist in solution as stable dimers (*24*). Moreover, the domains of both proteins critical for dimerization map to regions coincident with DNA binding, which includes the periodic leucine repeats. The possibility that a stable dimerization interface can be generated from an α helix of less than ten turns has been demonstrated in model systems with synthetic peptides (*25*).

If our "leucine zipper" model is correct, then the α helices must interlock in one of two orientations, parallel or anti-parallel relative to the amino-to-carboxyl dipole of each helix. Crystallographic studies have shown that helix packing can occur in either orientation (*17*). However, all precedents for coiled-coil interaction between two polypeptides adopt a parallel conformation (*20*). Despite this fact, we offer several reasons to anticipate that the "leucine zipper" motif exists in an antiparallel conformation. First, the intimate degree of side chain packing implicit to our model might only be compatible with antiparallel helices. Amino acid side chains are disposed at an angle pointing toward the amino terminus of the helix (Fig. 4); as such, an antiparallel arrangement might be better suited for side chain interlocking. Second, an antiparallel conformation would allow the dipole of one helix to attract, rather than repel, the dipole of the matching helix.

We anticipate that the issue of helical orientation will be central to the mode by which these proteins bind DNA. Either conformation, parallel or antiparallel, results in a rotationally symmetric molecule (Fig. 5B). In an antiparallel conformation, the axis of rotation would be perpendicular to the paired helices, whereas in a parallel conformation, the axis would be in line with the helices. The relevance of helix orientation is underscored by the fact that the "leucine zipper" alone is not sufficient to confer sequence-specific interaction with DNA. Amino acid sequence analysis of C/EBP has shown that the protein contains a high proportion of basic residues in a 30-amino-acid region immediately adjacent to its "leucine zipper" (*11*). This highly basic region of C/EBP exhibits substantial sequence similarity to a region of the Fos-transforming protein that is juxtaposed identically to its "leucine zipper" (*11*). Moreover, the C/EBP basic region must remain intact in order for C/EBP to bind DNA (*26*). We predict that the "leucine zipper" juxtaposes the basic regions of two polypeptides in a manner suitable for sequence-specific recognition of DNA (Fig. 6).

By comparing the amino acid sequences of several DNA binding proteins we have discovered a repeating motif of leucine residues. We propose that these leucines project from comparatively long, stable, solvent-exposed α helices, and that the leucine residues that

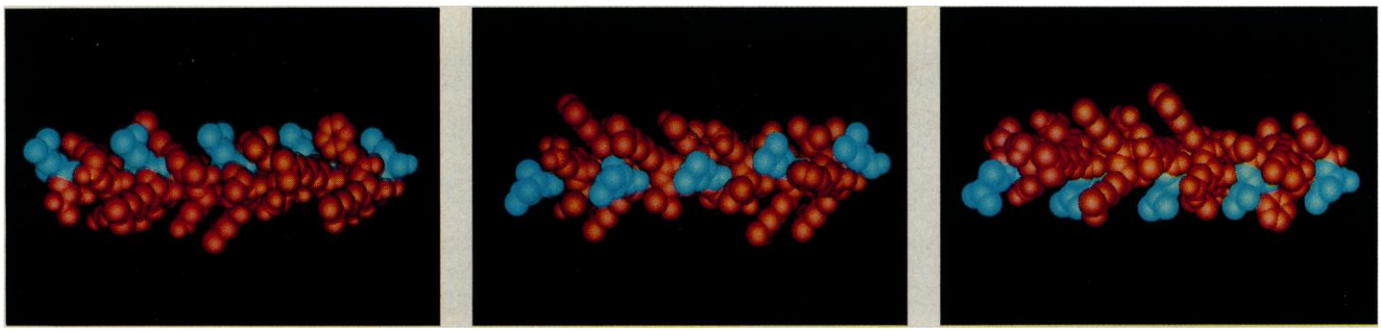


**Fig. 4.** Three-dimensional model of the helical leucine repeat region of C/EBP. The amino acid side chains of C/EBP between leucine 315 and leucine 343 were appended onto the α carbon backbone of an idealized α helix and displayed in three dimensions by means of a computer graphics program (*23*). Leucine residues that occur at every seventh residue are displayed in blue, all other residues are displayed in red. The amino-to-carboxyl dipole of the helix is arranged left to right. Each successive view of the helix is rotated by 90° around an axis parallel to the helix. Note that the distance between leucine residues exceeds the van der Waals radii of the side chain R groups and that side chains tilt toward the amino terminus of the helix.

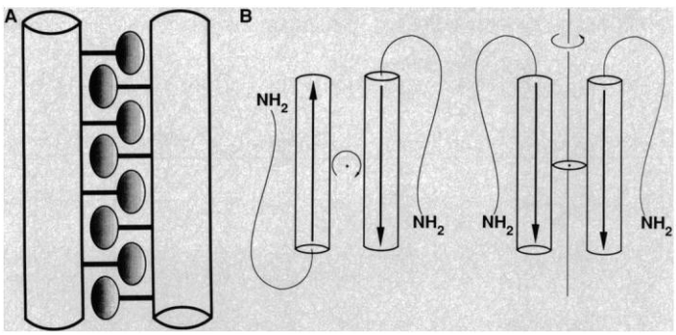



**Fig. 5.** Schematic diagram showing hypothetical interdigitation of leucine side chains between two α helices. (**A**) Two parallel tubes represent the approximate dimensions of the α carbon backbone of idealized α helices. Interdigitating protrusions symbolize leucine side chains. The sphere located at the tip of each residue represents the two methyl groups attached to the γ carbon atom of the leucine side chain. (**B**) C/EBP dimers disposed in either an antiparallel (left) or parallel array (right) with respect to the leucine repeat. Thick arrows within helical cylinders denote amino-to-carboxyl dipole. Thin arrows denote axis of rotational symmetry.

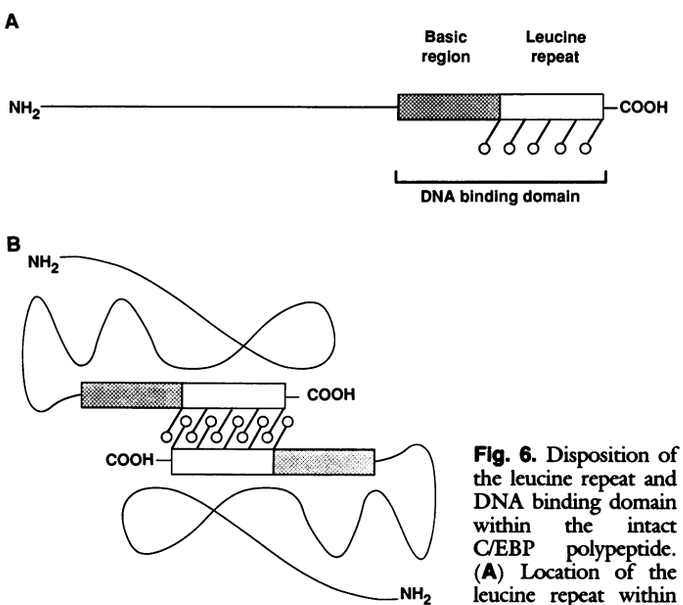



**Fig. 6.** Disposition of the leucine repeat and DNA binding domain within the intact C/EBP polypeptide. (**A**) Location of the leucine repeat within the intact C/EBP polypeptide. The region necessary for sequence-specific interaction with DNA extends beyond the leucine repeat toward the amino-terminus of the protein, and includes a 30-amino-acid region that is highly positively charged (stippled). (**B**) A hypothetical model of a C/EBP dimer established by the interdigitation of leucine repeat helices in an antiparallel conformation.

project from one α helix interdigitate with those of a second helix, causing the two molecules to dimerize. We believe that this motif, the "leucine zipper," represents a part of the scaffold that molds a protein to interact with its target site on DNA. Finally, we point out the possibility that the "leucine zipper" might allow dimerization of different polypeptides so long as each subunit contained the motif. Indeed, recent evidence has raised the possibility that the Fos and Jun transforming proteins, which we have shown to contain the leucine repeat motif, form a heterotypic complex (*27*). The notion of heterotypic interactions raises potentially important implications relating to the combinatorial action of gene regulatory proteins, and may facilitate a more direct attack on the function of nuclear oncogenes.

## REFERENCES AND NOTES


1. W. F. Anderson, D. H. Ohlendorf, Y. Takeda, B. W. Mathews, *Nature* **290**, 754 (1981); D. B. McKay and T. A. Steitz, *Nature ibid.*, p. 744; C. O. Pabo and M. Lewis, *Nature* **298**, 443 (1982); R. W. Schevitz, Z. Otwinowski, A. Joachimiak, C. L. Lawson, P. B. Sigler, *ibid.* **317**, 782 (1985).
2. T. A. Steitz, D. H. Ohlendorf, D. B. McKay, W. F. Anderson, B. W. Mathews, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3097 (1982).
3. D. R. Engelke, S.-Y. Ng, B. S. Shastry, R. G. Roeder, *Cell* **19**, 717 (1980); H. R. B. Pelham, and D. D. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4170 (1980); J. S. Hanas, D. J. Hazuda, D. F. Bogenhagen, F. Y.-H. Wu, C.-W. Wu, *J. Biol. Chem.* **258**, 14120 (1983).
4. R. S. Brown, C. Sander, P. Argos, *FEBS Lett.* **186**, 271 (1985); J. Miller, A. D. McLachlan, A. Klug, *EMBO J.* **4**, 1609 (1985).
5. G. P. Diakun, L. Fairall, A. Klug, *Nature* **18**, 698 (1986); A. D. Frankel, J. M. Berg, C. O. Pabo, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4841 (1987).
6. J. M. Berg, *Science* **232**, 485 (1986).
7. T. A. Hartshorne, H. Blumberg, E. T. Young, *Nature* **320**, 283 (1986); U. B. Rosenberg *et al.*, *ibid.* **319**, 336 (1986); J. T. Kadonaga, K. R. Carner, F. R. Masiarz, R. Tjian, *Cell* **51**, 1079 (1987); D. C. Page *et al.*, *ibid.*, p. 109; R. M. Evans and S. M. Hollenberg, *ibid.* **52**, 1 (1988).
8. C. O. Pabo and R. T. Sauer, *Annu. Rev. Biochem.* **53**, 293 (1984).
9. J. A. McClarin *et al.*, *Science* **234**, 1526 (1986).
10. I. Hope and K. Struhl, *Cell* **46**, 885 (1986); J. L. Pinkham, J. T. Oleson, L. P. Guarente, *Mol. Cell. Biol.* **7**, 578 (1987).
11. W. H. Landschulz, P. F. Johnson, E. Adashi, B. J. Graves, S. L. McKnight, *Genes Dev.*, in press.
12. R. E. Kingston, A. S. Baldwin, Jr., P. A. Sharp, *Nature* **312**, 280 (1984); C. Setoyama, R. Frunzio, G. Liau, M. Mudryi, B. de Crombrugghe, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3213 (1986); R. Kaddurah-Daouk, J. M. Green, A. Baldwin, Jr., R. Kingston, *Genes Dev.* **1**, 347 (1987); R. J. Distel, H.-S. Ro, B. S. Rosen, D. L. Groves, B. M. Spiegelman, *Cell* **49**, 835 (1987); K. Lech, K. Anderson, R. Brent, *ibid.* **52**, 179 (1988).
13. B. J. Graves, P. F. Johnson, S. L. McKnight, *Cell* **44**, 565 (1986); P. F. Johnson, W. H. Landschulz, B. J. Graves, S. L. McKnight, *Genes Dev.* **1**, 133 (1987).
14. P. K. Vogt, T. J. Bos, R. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3316 (1987).
15. K. Struhl, *Cell* **50**, 841 (1987).
16. G. E. Schulz and R. H. Schirmer, *Principles of Protein Structure* (Springer-Verlag, New York, 1979).
17. C. Chothia, *Annu. Rev. Biochem.* **53**, 537 (1984).
18. N. Sundaralingam, Y. C. Sekharudu, N. Yathindra, V. Ravichandran, *Proteins* **2**, 64 (1987).

19. S. Marqusee and R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8898 (1988).
20. F. H. C. Crick, *Nature* **170**, 882 (1952); D. A. D. Parry, W. G. Crewther, R. D. B. Fraser, T. P. MacRae, *J. Mol. Biol.* **113**, 449 (1977); F. D. McKeon, M. W. Kirschner, D. Caput, *Nature* **319**, 463 (1986); C. Cohen and K. C. Holmes, *J. Mol. Biol.* **6**, 423 (1963).
21. By ion pair we refer simply to the $i+3$ and $i+4$ juxtaposition of basic (lysine and arginine) and acidic (glutamate and aspartate) amino acids. Marqusee and Baldwin (*19*) have indicated that the $i+4$ glutamate-lysine arrangement fosters helix stability better than the $i+3$ arrangement, and that the stability imparted by pairs of oppositely charged residues is probably mediated by salt bridges.
22. M. Sundaralingam, W. Drendel, M. Greaser, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7944 (1985).
23. Computer graphic modeling was done with the Promodel software package (New England Biographics). We placed the side chains found in the leucine repeat region of C/EBP on the idealized α-helical backbone specified by the software. The helix maintains 3.6 residues per turn with uniformly constant phi and psi angles, helical pitch, and rise. Although the program uses the lowest average conformation for the side chains, it neither alters side chain conformation due to neighboring residues nor evaluates plausibility of a proposed structure; no steric hindrance for any residues was observed.
24. I. Hope and K. Struhl, *EMBO J.* **6**, 2781 (1987); W. H. Landschulz, P. B. Sigler, S. L. McKnight, unpublished data. The carboxyl-terminal 18 kD of C/EBP was expressed in bacteria and purified to homogeneity. A 1 $\mu M$ solution of the protein was treated with 0.001 percent glutaraldehyde at room temperature for 1 to 3 hours, and the products were analyzed by SDS–polyacrylamide gel electrophoresis. Within 3 hours, more than 90 percent of the 18-kD species was cross-linked into a 35-kD species.
25. S. Y. M. Lau, A. K. Taneja, R. S. Hodges, *J. Biol. Chem.* **259**, 13253 (1984).
26. W. H. Landschulz and S. L. McKnight, unpublished data. A recombinant DNA clone encoding C/EBP was mutagenized in vitro and expressed in bacteria. Amino-to-carboxyl deletion mutants were capable of sequence-specific interaction with C/EBP binding sites on DNA so long as the leucine repeat and adjacent basic region remained intact.
27. B. R. Franza, Jr., F. J. Rauscher III, S. F. Josephs, T. Curran, *Science* **239**, 1150 (1988); F. J. Rausher III *et al.*, *ibid.* **240**, 1010 (1988); R. Tjian and R. Turner, personal communication.
28. L. W. Stanton, P. D. Pahrlander, P. M. Tesser, K. B. Marcu, *Nature* **310**, 423 (1984).
29. F. Kaye *et al.*, *Mol. Cell. Biol.* **8**, 186 (1988); R. A. DePinho, D. S. Hatton, A. Tesfaye, G. D. Yancopoulos, F. W. Alt, *Genes Dev.* **1**, 311 (1987).
30. Y. Maki, T. J. Bos, C. Davis, M. Starbuck, P. K. Vogt, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2848 (1987).
31. C. van Beveren, F. van Straaten, T. Curran, R. Mueller, I. M. Verma, *Cell* **32**, 1241 (1983).
32. A. G. Hinnebusch, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6442 (1984).
33. We thank R. Kretsinger and P. Sigler for invaluable input; H. Smith for assistance with computer graphics; K. LaMarco for comments on the manuscript; C. Jewel for help with preparation of figures; C. Pratt for clerical assistance; R. Kingsbury for technical assistance; and C. Pabo, R. Baldwin, K. Yamamoto, A. Klug, and J. Berg for critical comment. Supported by a Medical Scientist Scholarship of the Life and Health Insurance Medical Research Fund (W.H.L.), postdoctoral fellowship awards from the Damon Runyon/Walter Winchell Cancer Fund and the Carnegie Corporation (P.F.J.), a Faculty Research Award from the American Cancer Society (S.L.M.), and grant award (S.L.M.) from NIH.

22 March 1988; accepted 20 May 1988


***Normally, of course, Edward wouldn't harm a fly.***