Scissors-Grip Model for DNA Recognition by a Family of Leucine Zipper Proteins

CHARLES R. VINSON, PAUL B. SIGLER, STEVEN L. MCKNIGHT

C/EBP is a sequence-specific DNA binding protein that regulates gene expression in certain mammalian cells. The region of the C/EBP polypeptide required for specific recognition of DNA is related in amino acid sequence to other regulatory proteins, including the Fos and Jun transforming proteins. It has been proposed that these proteins bind DNA via a bipartite structural motif, consisting of a dimerization interface termed the "leucine zipper" and a DNA contact surface termed the "basic region." An evaluation of the properties of conserved amino acids within the basic region of 11 deduced protein sequences, coupled with the observation that they are located at an invariant distance from the leucine zipper, has led to the formulation of a "scissors-grip" model for DNA binding. The architectural features of this model are well suited for interaction with directly abutted, dyad-symmetric DNA sequences. Data supportive of the model were obtained with chemical probes of protein:DNA complexes.

DIMERIZATION HAS EMERGED AS A UNIFYING PROPERTY of sequence-specific DNA binding proteins. The utility of dimerization can be understood to reflect the fact that the affinity of interacting molecules is exponentially related to the free energy of binding. If, excluding countervening effects, it can be assumed that a dimer acts as a single structural element, then it should double the DNA contact area and thereby square the affinity constant. Thus, tightly associated dimers are able to bind selectively to their DNA targets at far more dilute concentrations than monomers.

Duplex DNA, because of its relatively uniform conformation, would not seem well suited for generating specific binding surfaces. However, the special virtue of dimerization in DNA binding comes from the dyad symmetry of the DNA sugar-phosphate backbones. One need only to impose dyad symmetry on a local group of base pairs to generate a highly specific binding surface for a symmetrical, dimeric protein. For example, the inverted repeats that are recognized by type II restriction enzymes consist of as little as two or three base pairs per abutted half-site. Indeed, as exemplified by the operators recognized by prototypical bacterial repressors, dyad-related recognition elements can form avid binding sites even when separated by short segments of intervening DNA.

Binding sites on DNA impose several constraints on the proteins that must recognize them. (i) The DNA contact surface of a protein must be capable of penetrating either the major or minor groove of DNA in order to interact with the functional groups that distinguish one binding site from another. (ii) Dyad-symmetric binding sites further demand that the two contact surfaces of the binding protein be assembled so that they can simultaneously interact with both halves of the recognition site. The importance of this latter constraint can be appreciated if we realize that separating the two halves of a dyad binding site by an extra base pair causes their recognition surfaces to be displaced by a twist angle of 34.5° and an axial translation of 3.25 Å.

Two classes of DNA binding proteins have been analyzed in sufficient detail to demonstrate how the aforementioned constraints are met. Bacterial repressors of the helix-turn-helix category use a DNA contact surface composed mainly of two successive α helices arranged at nearly right angles. Dimerization results in a rotationally symmetric molecule wherein the DNA contact surface of one subunit is aligned antiparallel to that of the other subunit, and separated by approximately 34 Å. This conformation tailors protein subunits for interaction with symmetrical half-sites located on the same face of DNA, one helical turn apart (*1*). Dimers of the restriction enzyme, Eco RI, are also rotationally symmetric. In this second case, however, a "double-barreled helix" motif is used to project a tandem pair of α helices into the major groove of DNA on each half of the recognition site (*2*). The close apposition of symmetric DNA contact surfaces in the protein matches the direct abutment of half sites in the Eco RI recognition site.

The aforementioned paradigms, which have been realized from x-ray diffraction analyses of crystallized DNA:protein complexes, bear two important similarities. First, the protein moiety of each class adapts to the inverted repeat nature of its recognition site by forming a rotationally symmetric dimer. Second, α-helical structure is used by both helix-turn-helix proteins and Eco RI to either form or present the DNA contact surface. Guided by these and other basic principles of protein structure, we have developed a provisional model for a newly defined class of sequence-specific DNA binding proteins whose subunits contain a characteristic dimerization interface and a highly basic DNA contact surface that just precedes it. These proteins recognize binding sites on DNA that consist of directly abutted, dyad-symmetric half-sites. The model that we have developed proposes that two polypeptide chains join to form a Y-shaped molecule. The stem of the Y corresponds to a coiled pair of α helices. The bifurcating arms of the Y constitute a linked set of DNA

C. R. Vinson and S. L. McKnight are at the Howard Hughes Research Laboratories, Department of Embryology, Carnegie Institution of Washington, 115 West University Parkway, Baltimore, MD 21210. P. B. Sigler is at the Department of Molecular Biophysics and Biochemistry, Howard Hughes Medical Institute, Yale University, 260 Whitney Avenue, New Haven, CT 06511.

contact surfaces and emerge from the point at which the paired α helices separate. The bifurcation point is situated on the rotational axis of the dimer and contacts the major groove of DNA at the very center of the dyad-symmetric recognition site. Since the bifurcation point closely approaches the center of the DNA target, each arm of the Y is optimally positioned to begin tracking along each half of the recognition site (*3*).

The family of proteins considered below contains a DNA binding domain that consists of two regions (*4*). One region facilitates dimer formation and the other contacts DNA. The dimer-forming region, termed the "leucine zipper," is hypothesized to consist of two amphipathic α helices, about 30 amino acids long, that adhere tightly to one another through a narrow hydrophobic spine punctuated every seven residues by a leucine side chain. Considerable evidence substantiating this scheme is now available (*5–7*). The heptad repeat of leucine residues, together with disulfide cross-linking experiments (*6*), spectral studies (*6, 8*), and mutational studies (*7*), imply that the interacting α helices adhere to one another in parallel, forming a coiled-coil (*9*).

Agre *et al.* (*10*) have carried out in vitro binding studies with chimeric proteins that interchange the leucine zipper of one protein with the DNA contact region of another protein. These studies show that the region responsible for specific DNA binding is immediately NH_2-terminal to the leucine zipper. This region has a characteristic sequence rich in basic amino acid residues and has thus been termed the "basic region." We propose that the basic region generates the DNA contact surface at the bifurcation point of two zippered helices. This model is described below, and we refer to proteins that show the contiguous basic region–leucine zipper architecture as bZIP proteins to distinguish them from other transcriptional regulatory proteins that do contain a leucine zipper, yet whose DNA contact surface is not contiguous (*11*).

Although the role of the leucine zipper and its properties is beginning to be clearly defined, previous studies have not addressed the conformation of the basic region, nor have they considered the nature of its orientation relative to the adjacent, zippered α helices. As a first step in examining the properties of the basic region, we compared the amino acid sequences of 11 proteins that show the bZIP motif (Fig. 1). This list includes three plant proteins, four mammalian proteins, and four fungal proteins. Similarities have already been noted in the amino acid sequences of the basic regions and leucine zippers of bZIP proteins (*12*). What was not recognized in earlier studies was the conservation of exact spatial register between the basic region and the leucine zipper. All 11 proteins exhibit an invariant alignment between the basic region and leucine zipper (Fig. 1). The basic regions exhibit a consensus, 16-residue sequence that starts exactly *7* residues NH_2-terminal to the "first" leucine of the zipper (*13*).

The precise alignment of conserved residues between the basic region and zipper suggests that DNA binding depends on a stringently fixed three-dimensional relation between the dimerization interface and DNA contact surface. Indeed, when the register between the basic region and zipper is altered by either the insertion or deletion of five amino acids, sequence-specific recognition of DNA is eliminated (*10*).

Having recognized the importance of fixed spatial register, we considered the possibility that this register might reflect continued α helical structure progressing from the zipper into the basic region. Although no biophysical studies have been conducted to examine the structural properties of the basic region of bZIP proteins, we offer two observations that are consistent with prediction that it will exist, to a large extent, as α helix. First, the basic regions of all 11 proteins examined (Fig. 1) are completely free of helix-destabilizing residues (proline and glycine). Moreover, the six–amino acid region that links the basic region to the zipper is also free of proline and glycine in all 11 proteins. Adding up the 16 residues of each basic region, and the 6 residues of each "linker" region, a total of 244 positions are free of helix-destabilizing residues. The second observation consistent with α-helical structure is the occurrence of oppositely charged amino acids (basic and acidic residues) configured in an arrangement three or four amino acids apart. The frequent occurrence of this arrangement of charged amino acids, which was also noted in the case of the C/EBP leucine zipper (*4*), is known to correlate with helix stability (*14*).

Persistence of α-helical structure NH_2-terminal to the zipper would not preserve the amphipathic "phase" of the zipper. Instead of juxtaposing attractive, hydrophobic residues suitable for a dimerization interface, the α-helical surfaces of the basic region seem designed to repel one another. For example, if the coiled-coil of the zipper persisted into the basic region, the side chains located seven residues NH_2-terminal to the first leucine would face their symmetrical counterpart. Since at this point the opposing residues are almost always arginine or lysine (Fig. 1), it is likely that zippered α helices would disengage and bifurcate like the arms of a Y.

We predict that the bifurcation point located at the junction of the zipper and basic region will specify the formation of two, tightly

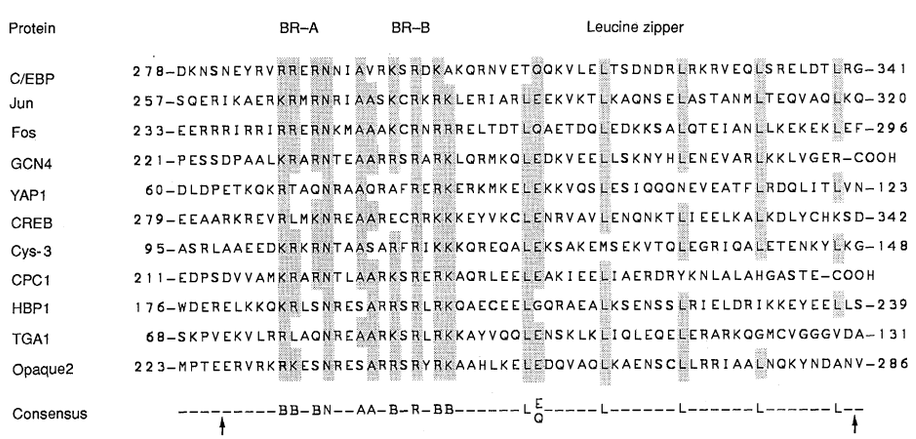

Fig. 1. Amino acid sequences corresponding to the DNA binding domains of 11 bZIP proteins. Protein identities are designated in the left hand column (*25*). Numbers preceding each sequence correspond to the distance, in amino acid residues, from the NH_2-terminus of the respective protein. For example, C/EBP contains 278 residues NH_2-terminal to the aspartic acid (D) displayed at the beginning of its amino acid sequence. Numbers following each sequence correspond to the next amino acid in the intact protein. Two proteins, GCN4 and CPC1, terminate within the displayed sequence. Polypeptide sequences (*25*), designated according to the single letter code (*26*), were aligned according to the positions of two invariant amino acid residues. One invariant residue corresponded to the asparagine (N) located at amino acid position 292 of C/EBP; the other corresponded to the arginine (R) located at position 300 of C/EBP. The 11 proteins contain a cluster of similar residues in areas designated basic region A (BR-A) and basic region B (BR-B). This cluster of similar residues, designated as a consensus below the 11 protein sequences, is positioned at an invariant distance from the periodic repeat of leucines that define the leucine zipper. Arrows at the bottom left and right designate the limits of amino acid sequence included in the computer graphic model shown in Fig. 2. See (*25*) for information on the proteins.

linked DNA contact surfaces—each corresponding to an arm of the Y. Furthermore, the cleft of the bifurcation point should meet DNA at the center of a dyad-symmetric binding site, allowing the two arms of the Y to match the two halves of a dyad-symmetric recognition sequence. The close apposition of the two DNA contact surfaces predicted by this model fits well with the properties of bZIP binding sites. That is, the DNA targets for bZIP proteins are directly abutted, inverted repeats (*15, 16*).

To investigate this proposal in more detail, we built a "ball and stick" model, using canonical B-form DNA and α helices of standard dimensions (*17*). We found that it was possible to cause bifurcating α helices to begin tracking in opposite directions along the major groove of DNA. Subsequent refinement of the model with computer graphics showed that the backbone torsion angles were well within the limits of canonical α helices (*18*). When we positioned bifurcating helices into the major groove of DNA, certain consensus basic residues became naturally disposed toward the sugar-phosphate backbone. However, if the entire basic region was configured as continuous α helix, the bifurcating arms of the Y protruded from the back side of the double helix. In other words, the portion of the basic region located closest to the zipper (designated BR-B in Fig. 1) fit snugly into the major groove of DNA, whereas the more distal part (BR-A) extended out the back side of the double helix.

There are two reasons why both parts of the basic region (BR-A and BR-B) should come in close contact with DNA. First, both parts are highly positively charged. Second, mutations in either part block DNA binding (*7*). For reasons outlined above, we anticipate that the basic region is largely α helical. If BR-A is to continue to track in the major groove, the hypothetical helix must be broken at a point between BR-A and BR-B, thus allowing BR-A to be turned in a different direction. Thus, our attention became focused on the task of modeling a turn between BR-A and BR-B.

Searches for the nonrandom distribution of amino acids within solved crystal structures have shown that asparagine, when normalized to its overall abundance, is the amino acid that occurs most frequently at the NH_2-terminus of α helices (*19*). The suitability of asparagine for the NH_2-terminus of α helices derives from the fact that the oxygen of its side chain is capable of forming a hydrogen bond with the peptide nitrogen two or three residues into the α helix (*19, 20*). This interaction terminates the continuity of hydrogen bonds necessary for α helical structure, thereby "capping" the NH_2-terminus of the helix in a structure termed an N-cap. Noting that an asparagine residue occurs at an invariant position between BR-A and BR-B in all 11 bZIP proteins (Fig. 1), we introduced an N-cap at the terminus of the hypothetical helix that connects the zipper to BR-B by hydrogen bonding the γ oxygen of the invariant asparagine to the peptide nitrogen three residues into the helix.

By terminating the helix at a point between BR-A and BR-B, it was possible to place the helix including BR-A in a position that allowed its continued interaction with the major groove of DNA. This produced a 75° angle between successive α-helical axes. The backbone torsional angles of $\phi = -123°$ and $\psi = 178°$ for the invariant asparagine, and $\phi = 60°$ and $\psi = -150°$ for the preceding arginine, formed a turn typical of those generated by the N-caps cited by Richardson and Richardson (*19*). Moreover, positioning of the BR-A helix to track in the major groove placed the proximal methylenes of the arginine in close apposition to the methyl groups of the conserved alanines located three and four amino acids away (Fig. 1), thereby offering the opportunity for stabilizing interactions.

The use of an asparagine to angulate the basic region, and thereby maintain contact between BR-A and DNA, contributes two favorable features to the model. First, the nitrogen of the highly fixed amide function provides a potential hydrogen bond to the phosphodiester backbone of DNA. Second, asparagine does not neutralize, and thereby block, the potential interaction between the positive pole of the α-helical macrodipole (*21*) and a negatively charged phosphate. Neither feature would be provided by aspartic acid, a negatively charged amino acid with similar stereochemical features that also frequently forms an N-cap (*19*).

It is possible that the polypeptide segment corresponding to BR-A mediates close contact with DNA by use of a structural motif other than α helix. If, as assumed for the sake of simplicity, α-helical structure is adopted on the NH_2-terminal side of the hypothetical asparagine cap, it is likely to be disrupted shortly after passing through BR-A. This prediction rests on the observation that 5 of the 11 bZIP proteins contain a proline residue within 8 residues of BR-A (Fig. 1). Finally, even if the polypeptide segment corresponding to BR-A is capable of forming α-helical structure and adopting a 75° bend relative to the BR-B helix, such conformation is unlikely to be rigidly formed in the absence of DNA, as this would prematurely "close the clamp" of the protein's DNA binding surface (Fig. 2).

A hypothetical dimer of a bZIP protein is shown bound to B-form DNA in Fig. 2. The polypeptide chains have been assembled according to the canonical dimensions of coiled-coil α helices (*18*). Paired helices are shown bifurcating within the six–amino acid segment that links the leucine zipper to the basic region. Helical structure is shown persisting along each subunit through the sequence corresponding to BR-B, with one subunit tracking in an upward direction along the major groove (in front of the DNA) and the other subunit tracking in an opposite direction, down the major

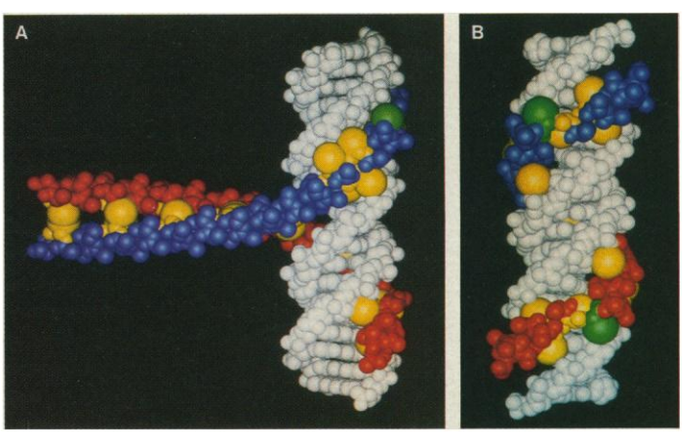

Fig. 2. Hypothetical model of a dimeric bZIP protein interacting with its target DNA. Straight, helically uniform DNA is shown as a light gray CPK model with standard van der Waals radii. The 20-bp sequence, 5′-TGCAGATTGCGCAATCTGCA-3′, contains at its center a high-affinity C/EBP binding site. Precisely the same 20-bp sequence was used in the methylation protection and hydroxyl radical footprinting studies (Figs. 3 to 5). The polypeptide sequence that was modeled corresponds to a 57–amino acid sequence progressing from asparagine residue 283 of C/EBP to arginine residue 339 (Fig. 1). Side chains are represented by CPK alanines with the exception of consensus residues. One subunit is displayed in red and the other is blue. Helices of the zipper are arranged in a canonical coiled-coil conformation (*18*). The leucines of the zipper and the consensus basic residues are yellow, the invariant asparagine (cap) is green. The side chains of the consensus basic residues are represented by an enlarged γ carbon to show how they are disposed in relation to the DNA and coiled-coil. (**A**) A "side view" in which the leucine zipper is to the left, and the axis of the coiled-coil is horizontal and parallel to the plane of the paper. (**B**) A "backview" in which the most NH_2-terminal element of the DNA binding domain (BR-A) continues to track in the major groove of DNA after a sharp (75°) bend at the asparagine cap (green). A rotation of 180° around the center of the panel gives the same structure. That is, there is a twofold axis of rotational symmetry, or dyad, that is perpendicular to the plane of the paper. This dyad passes through the center of the symmetrical binding site and down through the dimerization interface of the leucine zipper (obscured by the DNA).

groove (behind the DNA). According to this model, the bZIP dimer begins its contact with DNA at the point where the dyad halves of the DNA recognition site abut one another. This is in sharp contrast to the way helix-turn-helix proteins bind to successive major grooves one helical turn apart.

An additional feature of the model (Fig. 2) concerns the disposition of highly conserved basic residues in BR-A and BR-B. The positively charged side chains of conserved basic residues are naturally disposed toward the edges of the major groove where they might interact with the negatively charged phosphodiester backbone. One such example is the invariant arginine residue located in the approximate center of BR-B (Fig. 1). As the helix including BR-B tracks along the major groove, this arginine residue comes in close contact with the phosphodiester backbone between two and three residues from the center of the dyad. The position of this invariant arginine is best visualized in Fig. 2B, where its side chain can be located as a yellow sphere immediately above the invariant asparagine (green) of one subunit (red) and below the invariant asparagine of the other subunit (blue).

Thus far we have outlined a series of observations that provide a simplified architectural model of how bZIP proteins might bind to DNA. The model predicts that a rotationally symmetric dimer will form as a result of the leucine zipper and that the DNA contact surface will be formed of bifurcating, NH_2-terminal extensions of the zipper's α helices. The two helices would then track along the major groove of DNA in opposite directions, corresponding to the dyad-symmetric nature of DNA recognition sites common to bZIP proteins. Finally, we hypothesize that there is a characteristic point where the presumably continuous DNA-contact helices of bZIP proteins become broken and angulated. The formation of this bend, which perhaps occurs only in the presence of DNA, allows bZIP proteins to "grip" around the DNA on the side opposite to their initial approach. We liken this arrangement to the "scissors grip" that a wrestler uses to grasp the torso of his opponent. This constitutes an unusually simple DNA binding motif that relies almost exclusively on α-helical structure.

High-resolution structural studies of bZIP proteins would provide the most rigorous test of the model presented above. Neither x-ray diffraction nor two-dimensional nuclear magnetic resonance studies have been completed on this class of DNA binding proteins. We have, however, conducted methylation protection and hydroxyl radical footprinting analyses, which provide preliminary tests of three predictions of the "scissors-grip" model.

An obvious prediction of the model presented in Fig. 2 is that bZIP proteins will establish symmetrical contacts with substrate DNA. The data in Figs. 3, 4, and 5 support this expectation. Occupancy of a symmetric binding site by C/EBP alters the pattern of methylation by dimethyl sulfate (DMS) at four locations on each DNA strand. Bound protein partially blocks the methylation of two guanine residues and accentuates methylation of one adenine residue and one guanine residue (Fig. 3, left). The same pattern of protection and enhancement is observed on both DNA strands (Fig. 5). Moreover, these patterns are symmetric relative to the half-sites of the recognition site.

Evidence of symmetrical interaction was also observed in the pattern of hydroxyl radical cleavage of an occupied C/EBP binding

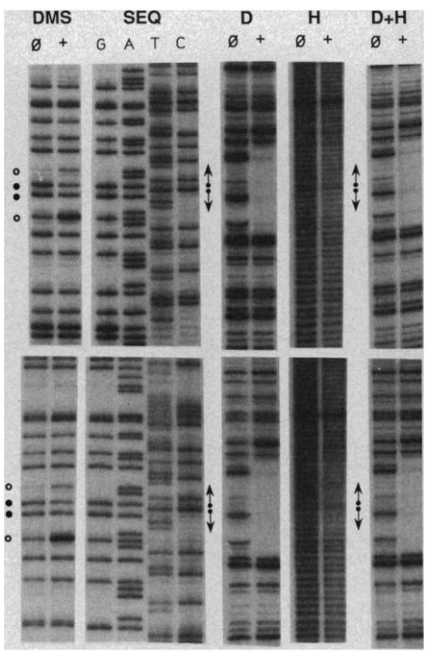

Fig. 3. Effects of C/EBP binding on methylation by dimethyl sulfate, and cleavage by DNase I and hydroxyl radical. Left panels (DMS) show patterns of methylation protection and enhancement resulting from C/EBP interaction with DNA. DNA containing a high-affinity C/EBP binding site was terminally labeled with radioactive phosphate and exposed to dimethyl sulfate either in the absence (ϕ) or presence (+) of C/EBP protein. Bands designated by a filled circle represent positions on the DNA that were methylated less effectively in the presence of C/EBP (methylation protection), bands designated by an open circle represent positions that were methylated more effectively in the presence of C/EBP (methylation enhancement). Protected and enhanced bands were identified by chemical sequencing of the same DNA molecules (SEQ). Diverging arrows show the location of the dyad symmetric C/EBP binding site (5′-ATTGCGCAAT-3′). DNase I and hydroxyl radical footprints were also carried out on the same, radioactively labeled DNA molecules. The panels labeled (D) show DNase I footprints established by C/EBP, and the panels labeled (H) show hydroxyl radical footprints. (ϕ) and (+) refer to cleavage reactions carried out in either the absence or presence of C/EBP. Panels (D+H) show DNase I and hydroxyl radical footprints carried out concomitantly. The length of time of autoradiographic exposure used to generate the images shown in panels (D) and (D+H) were one-tenth that used to generate the images shown in panel (H). Longer exposure of panels (D+H) showed evidence of hydroxyl radical cleavage as a background to DNase I cleavage. Upper panels represent experiments carried out with DNA labeled on the opposite strand from those of the lower panels (*27*).

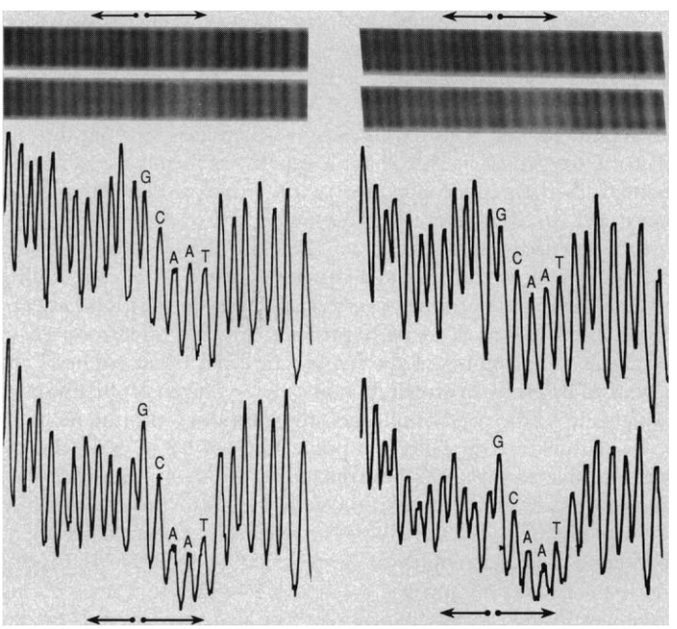

Fig. 4. Densitometric scans of hydroxyl radical cleavage observed in the absence and presence of DNA-bound C/EBP. Upper panels show autoradiographic exposures corresponding to hydroxyl radical cleavage patterns generated in either the absence (top exposure) or presence (bottom exposure) of C/EBP. Left panel shows cleavage pattern on opposite DNA strand from right panel (*27*). The position of the dyad-symmetric C/EBP binding site is indicated by diverging arrows and was established by alignment of hydroxyl radical cleavage products with chemically sequenced DNA that had been subjected to electrophoresis on the same polyacrylamide gel. Each exposure was scanned by densitometry so that peak heights could be evaluated quantitatively (Fig. 5). Individual residues are identified on one half of each dyad (5′-GCAAT-3′).

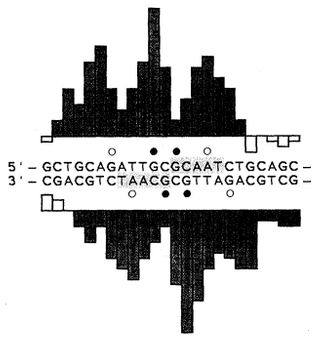

Fig. 5. Summary of effects of C/EBP on methylation by dimethyl sulfate and cleavage by hydroxyl radical. DNA sequence is stippled in regions corresponding to both halves of the dyad symmetric C/EBP binding site (5′-GCAAT-3′). Open circles directly above and below the DNA sequence identify residues that are methylated more readily in the presence of DNA-bound C/EBP. Closed circles identify residues that are protected from methylation by bound protein. Histograms above and below the DNA sequence identify residues that were cleaved by hydroxyl radical either less effectively (stippled bars) or more effectively (open bars) in the presence of DNA-bound C/EBP.

site; C/EBP produces a modest hydroxyl radical footprint on the symmetric sequence 5′-ATTGCGCAAT-3′ (Figs. 3 and 4). We tested the possibility that the hydroxyl radical, or conditions used for its cleavage, might hamper C/EBP binding, thus diminishing the apparent protection by the protein. This was accomplished by carrying out a deoxyribonuclease I (DNase I) footprint in the presence of the hydroxyl radical probe. As shown in Fig. 3 (right), the C/EBP binding site was fully protected from DNase I cleavage throughout the 1-minute exposure to the hydroxyl radical. We interpret the lightness of the hydroxyl radical footprint to reflect the fact that bZIP proteins interact exclusively with the major groove of DNA. As shown in Fig. 2, bZIP dimers are predicted to begin DNA contact in the major groove near the center of their dyad-symmetric binding sites and never "cross over" a single phosphodiester backbone. Since the hydroxyl radical cleaves DNA via reactions initiated predominately within the minor groove of DNA (*22*), binding of bZIP proteins should not result in an effective steric block to this chemical probe.

Regardless of the modest protection from hydroxyl radical cleavage afforded by C/EBP, it was possible to map residues within and around the binding site where cleavage was affected by bound C/EBP. Autoradiographic exposures (upper panels of Fig. 4) were scanned with a densitometer (lower panels of Fig. 4). Absorbance peaks were assigned to specific residues within the dyad symmetric binding site by comparison with chemically sequenced DNA that had been sized on the same electrophoresis gel. Comparison of peak heights for each residue (Fig. 5) reveals symmetrical patterns of hydroxyl radical protection. Three matching areas of protection occur on each DNA strand.

Two additional observations that emerged from hydroxyl radical footprinting are supportive of our model (Fig. 2). First, the regions of DNA protected from hydroxyl radical attack match the regions that the model predicts to be contacted by bZIP proteins. Although the dyad symmetric recognition sequence for C/EBP encompasses only ten residues, the scissors-grip model predicts intimate interaction covering 16 bp of DNA—just the span of DNA protected from hydroxyl radical cleavage. The second aspect of the hydroxyl radical footprint pattern that is consistent with our model is the stagger of protection. That is, bound C/EBP leads to a pattern of protection that extends on each DNA strand two or three bases farther from the dyad center in a 5′ direction than in a 3′ direction. This "5′ stagger" is consistent with major groove interaction and precisely opposite from the protection stagger known to result from minor groove interaction (*23*).

The basis for 5′ stagger generated by a bZIP protein can be visualized in Fig. 2A. The phosphodiester contacts most distal to the center of the dyad are predicted to be established by the BR-A region of each subunit and occur on the side of DNA opposite from the bifurcation point of the leucine zipper. These phosphodiester bonds correspond to the trinucleotide sequence 5′-CAG-3′ located four residues from the respective 5′ termini of the sequence shown in Fig. 5. Protection from hydroxyl radical cleavage was indeed observed at this trinucleotide sequence on each DNA strand (Fig. 5). However, little or no protection was observed at the positions of the complementary residues on the other DNA strand. No protein contacts should occur on the phosphodiester backbone of the DNA strand opposite to the aforementioned 5′-CAG-3′ trinucleotide (Fig. 2A).

Although the scissors-grip model provides a plausible architectural scheme for stable interaction between protein and DNA, it raises the question as to how bZIP proteins engage and disengage from their substrates. In the absence of DNA, bZIP proteins are unlikely to exist in a rigidly defined conformation analogous to that shown in Fig. 2. For example, the sharply angulated bend within the basic region may be a dynamic structure stabilized only on interaction with the appropriate binding site on DNA. In the absence of DNA, the proposed α-helical segments of the basic region, and the 75° bend at the asparagine cap, are almost certainly less well defined than depicted, thus facilitating access to DNA. This view is analogous to that proposed for the enhanced α-helical stability of protamines that results on exposure to crystalline transfer RNA (*24*). Moreover, even under conditions where equilibrium favors formation of the protein-DNA complex, transient alternative conformations of the protein should exist. Some of these, such as momentary dissolution of the BR-A helix, or destabilization of the N-cap, may allow the DNA to "escape" the scissors grip.

We have outlined a model that offers a plausible scheme for the way in which a newly recognized class of regulatory proteins bind DNA. These proteins use the leucine zipper motif to form a closely juxtaposed set of DNA contact surfaces, each of which fits into half of a dyad-symmetric recognition sequence. The modeling relied heavily on paradigms established from crystallographic studies of other proteins, yet was undertaken in the absence of any high-resolution structural information. The feature of this hypothetical protein:DNA complex that lends itself to modeling is the presumed dominance of a single secondary structural element in each of the macromolecular partners, that is, B-DNA and α-helical polypeptides. Indeed, the proposed structure is devoid of tertiary interactions between residues within the same strand that are widely separated in the sequence. Even the sharp turn hypothesized at the asparagine cap is accomplished by interactions between neighboring residues. Instead, the structure appears to achieve its global stability from interactions between the four constituent strands—the two zippered α helices and the two complementary strands of B-DNA.

REFERENCES AND NOTES

1. A. K. Aggarwal, D. W. Rodgers, M. Drottar, M. Ptashne, S. C. Harrison, *Science* **242**, 899 (1988); S. R. Jordan and C. O. Pabo, *ibid.*, p. 893; Z. Otwinowski *et al.*, *Nature* **335**, 321 (1988).
2. J. A. McClarin *et al.*, *Science* **234**, 1526 (1986).
3. The architecture of this dyad-symmetrical interaction between protein and DNA is different from that of helix-turn-helix proteins, which use a pair of symmetrically disposed DNA contact surfaces to recognize dyad-symmetric half-sites that are separated by one full turn of duplex DNA. The model developed for DNA recognition by bZIP proteins predicts that two symmetrically disposed DNA contact surfaces emerge from a common point and, starting at the center of the dyad-symmetric DNA target site, straddle each half of the dyad. The geometrical basis for this difference is that, in helix-turn-helix proteins, the common dyad shared by the dimeric protein and symmetrical DNA target passes through the protein:DNA contact surface in the minor groove of DNA. The major grooves of the DNA present themselves to the protein a half-turn from the dyad (one helical turn apart from one another). In the case of bZIP proteins, the common dyad passes through the protein:DNA contact surface in the major groove. Thus, both the DNA contact surface of the protein and the symmetrical half-sites of the DNA target abut the symmetry axis.
4. W. H. Landschulz, P. F. Johnson, S. L. McKnight, *Science* **240**, 1759 (1988).
5. T. Kouzarides and E. Ziff, *Nature* **336**, 646 (1988); P. Sassone-Corsi, W. W. Lamph, M. Kamps, I. Verma, *ibid.*, p. 692.

6. E. K. O'Shea, R. Rutkowski, P. S. Kim, *Science* **243**, 538 (1989).
7. W. H. Landschulz, P. F. Johnson, S. L. McKnight, *ibid.*, p. 1681; R. Turner and R. Tjian, *ibid.*, p. 1689; R. Gentz, F. J. Rauscher III, C. Abate, T. Curran, *ibid.*, p. 1695; M. Neuberg, M. Schuermann, J. B. Hunter, R. Muller, *Nature* **338**, 589 (1989).
8. W. H. Landschulz, D. Shortle, S. L. McKnight, unpublished observations.
9. F. H. C. Crick, *Nature* **170**, 882 (1952).
10. P. Agre, P. F. Johnson, S. L. McKnight, *Science* **246**, 922 (1989).
11. P. F. Johnson and S. L. McKnight, *Annu. Rev. Biochem.* **58**, 799 (1989).
12. P. K. Vogt, T. J. Box, R. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3316 (1987); W. H. Landschulz, P. F. Johnson, E. Adashi, B. J. Graves, S. L. McKnight, *Genes Dev.* **2**, 786 (1988).
13. Of the 11 proteins (Fig. 1), 10 contain a leucine residue positioned seven residues away from the consensus sequence of the basic region. The single exception to this canonical arrangement is observed in the sequence of C/EBP, which contains a threonine residue in place of leucine. Although C/EBP does not exhibit the canonical pattern in this respect, its variant amino acid (threonine) is flanked by residues that frequently flank the "first" leucine of the other ten proteins. Ten of the 11 proteins exhibit either a glutamic acid or glutamine residue at the position immediately following the "first" leucine, and 7 exhibit one of these two residues at the position immediately preceding the "first" leucine. Although we are unaware of the significance of this arrangement of glutamic acid and glutamine residues, their presence at positions flanking the relevant threonine residue of C/EBP is consistent with the anticipation that C/EBP will, within this aspect of its DNA binding domain, employ the same general architecture as the other ten proteins.
14. N. Sundaralingam *et al.*, *Proteins* **2**, 64 (1987).
15. B. J. Graves, P. F. Johnson, S. L. McKnight, *Cell* **44**, 565 (1986).
16. K. Arndt and G. R. Fink, *Proc. Natl. Acad. Sci. U.S.A* **83**, 8516 (1986); D. E. Hill, I. A. Hope, J. P. Macke, K. Struhl, *Science* **234**, 451 (1986).
17. A molecular "ball and stick" model was constructed with atomic models (Maruzen Co. Ltd., Hinomoto Plastics, Tokyo, Japan).
18. The computer graphics model was built with an Evans and Sutherland PS390; we used the Rice version (6.6) of FRODO [T. A. Jones, *Computational Crystallography*, D. Sayre, Ed. (Clarendon, Oxford, 1982), pp. 303–317]. The starting point for the leucine zipper was energetically idealized coordinates for a polyalanine coiled-coil [D. A. D. Parry and E. Suzuki, *Biopolymers* 7, 189 (1969)], in which the methyl groups of alanines were replaced by the side chains of C/EBP. The α helices were separated to a center-to-center distance of 11.2 Å in order to visually accommodate the conserved leucine side chains in an all staggered conformation. There are two different relative azimuthal orientations of the opposing α helices in a coiled-coil that bring the leucine side chains into van der Waals contact. The one used here is referred to as the "4-3" repeat (*6*) and places both the leucine and typically hydrophobic side chain four residues carboxyl terminal to the leucine on the intersubunit contact surface. Coordinates for the target DNA were obtained by superimposing the base pairs of Fig. 5 on a canonical B-DNA backbone [S. Arnott and D. W. L. Hukins, *Biochem. Biophys. Res. Commun.* **47**, 1504 (1972)] with a program written by Z. Otwinowski. The model was built so that α-helical segments tracked snugly in the major grooves while disposing the consensus basic side chains toward the sugar-phosphate backbone. Although steric clashes and unfavorable backbone torsion angles were avoided, no attempt was made to minimize the internal energy of this model or otherwise compute its optimal structure. The computer files used to generate the graphic model shown in Fig. 2 are available on request.
19. J. S. Richardson and D. C. Richardson, *Science* **240**, 1648 (1988).
20. L. G Presta and G. D. Rose, *ibid.*, p. 1632.
21. W. G. J. Hol, P. T. van Duijnen, H. J. C. Berendsen, *Nature* **273**, 443 (1978).
22. T. Tullius, B. Dombroski, M. Churchill, L. Kan, *Methods Enzymol.* **155**, 537 (1987); A. M. Burkhoff and T. D. Tullius, *Cell* **48**, 935 (1987).
23. R. S. Youngquist and P. B. Dervan, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2565 (1985).
24. R. W. Warrant and S.-H. Kim, *Nature* **271**, 130 (1978).
25. Amino acid sequences included in Fig. 1 were obtained from the following sources: C/EBP, W. H. Landschulz, P. F. Johnson, E. Y. Adashi, B. J. Graves, S. L. McKnight, *Genes Dev.* **2**, 786 (1988); Jun, Y. Maki, T. J. Bos, C. Davis, M. Starbuck, P. K. Vogt, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2848 (1987); Fos, T. Curran, G. Peters, C. Van Beveren, N. M. Teich, I. M. Verma, *J. Virol.* **44**, 674 (1982); GCN4, A. Hinnebusch, *ibid.* **81**, 6442 (1984); G. Thireos, M. D. Penn, H. Greer, *ibid.*, p. 5096; YAP1, W. S. Moye-Rowley, K. D. Harshman, C. S. Parker, *Genes Dev.* **3**, 283 (1989); CREB, J. P. Hoeffler, T. E. Meyer, Y. Yun, J. L. Jameson, J. F. Habener, *Science* **242**, 1430 (1988); CYS3, Y.-H. Fu, J. V. Paietta, D. G. Mannix, G. A. Marzluf, *Mol. Cell. Biol.* **9**, 1120 (1989); CPC1, J. L. Paluh, M. J. Orbach, T. L. Legerton, C. Yanofsky, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3728 (1988); HBP1, T. Tabata *et al.*, *Science* **245**, 965 (1989); TGA1, F. Katagiri, E. Lam, N.-H. Chua, *Nature* **340**, 727 (1989); Opaque2, R. J. Schmidt, F. A. Burr, M. J. Aukerman, B. Burr, personal communication.
26. 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.
27. An oligonucleotide containing a high-affinity C/EBP binding site was synthesized (5′-GATCAAGCTGCAGATTGCGCAATCTGCAGCTT-3′), annealed, and inserted into a Bam HI restriction site that had been introduced into the herpes simplex virus thymidine kinase (*tk*) gene. The Bam HI site was located between 6 and 16 bp upstream from the *tk* mRNA start site, in a derivative of the *tk* promoter termed linker scanning mutant −16/−6 [S. L. McKnight and R. Kingsbury, *Science* **217**, 316 (1982)]. Two probe DNA molecules were prepared so that each DNA strand could be analyzed for methylation protection, DNase I footprinting, hydroxyl radical footprinting, and DNA sequencing. One probe was prepared by digesting the *tk* DNA at a Bgl II restriction site located 56 bp downstream from the *tk* mRNA start site, end-labeling with [γ-^{32}P]ATP, then digesting at an Eco RI site located 80 bp upstream from the mRNA start site. The other probe was prepared by the same methods but with the order of restriction endonuclease digestion reversed. Methylation protection and DNase I footprinting assays were as described [P. F. Johnson, W. H. Landschulz, B. J. Graves, S. L. McKnight, *Genes Dev.* **1**, 133 (1987)]. Hydroxyl radical footprinting assays were done by the methods of T. D. Tullius and B. A. Dombroski [*Proc. Natl. Acad. Sci. U.S.A.* **83**, 5469 (1986)]. C/EBP protein used in footprinting and methylation protection assays was an 88–amino acid fragment of the native polypeptide coterminal with the carboxyl terminus of the intact protein. This protein fragment was purified to homogeneity after overproduction in *Escherichia coli* (J. Shuman, unpublished results). Densitometric tracing of hydroxyl radical cleavage patterns was carried out with the use of a Joyce Loebl Chromoscan 3. Quantitation of differences in cleavage patterns in the absence and presence of bound protein shown in the histograms of Fig. 5 was accomplished by superimposing two tracings and taking direct measurements of differences in peak heights.
28. We thank R. Kretsinger, J. Berg, and C. Pabo for comments and guidance during the construction of the Maruzen "ball and stick" model; J. Shuman for providing purified C/EBP; Z. Otwinowski for help with computer modeling; P. Vogt for help in preparation of Fig. 2; J. Hayes and T. Tullius for advice on hydroxyl radical footprinting; R. Kingsbury for technical assistance; C. Norman for clerical help; K. LaMarco, M. Ptashne, and C. Emerson for comments on the manuscript; T. Tabata, R. Schmidt, B. Burr, F. Katagiri, and N.-H. Chua for unpublished amino acid sequences included in Fig. 1; and our many colleagues at the Carnegie Embryology Department for advice and encouragement. Supported by a postdoctoral fellowship from the American Cancer Society (C.R.V.) and by the Carnegie Institution of Washington and the Howard Hughes Medical Research Institute (S.L.M.).

31 July 1989; accepted 10 October 1989