merase chain reaction (PCR) amplification of pUC9-Sc4251 (*33*) with synthetic 5′-^{32}P–labeled or unlabeled 17-residue primers defining a 231-bp PCR product with the GRE centered. Nuclease digestion (90 s at 4°C) was initiated by addition of 0.2 μg DNase I (Sigma) and $CaCl_2$ to 2.5 mM and was quenched by addition of 200 μl of 1% SDS, 200 mM NaCl, 20 mM EDTA, and yeast transfer RNA (25 μg/ml) (Sigma). Samples were purified by phenol-chloroform extraction and ethanol precipitation and were run on a 6% sequencing (7.7 M urea) polyacrylamide gel.

27. α-Helices can be inferred from CD spectra obtained in aqueous solution with much higher confidence than for other secondary structures, although strictly speaking we cannot distinguish between α- and 3_{10}-helices; see R. W. Woody, in *The Peptides*, S. Undenfriend, J. Meienhofer, J. R. Hruby, Eds. (Academic Press, New York, 1985), vol. 7, pp. 15–114; Y.-H. Chen, J. T. Yang, K. H. Chau, *Biochemistry* **13**, 3350 (1974).
28. CD experiments suggest that the helical content in GCN4-bZIP1 also increases substantially upon binding GRE20 (A. D. Frankel, E. K. O'Shea, T. G. Oas, P. S. Kim, unpublished results). These experiments are difficult to interpret, however, because preliminary two-dimensional nuclear magnetic resonance studies (L. P. McIntosh, T. G. Oas, P. S. Kim, unpublished results) indicate that the leucine zipper region of GCN4-bZIP1 is substantially less stable than the isolated leucine zipper (*6*, *7*).
29. K. T. O'Neil, R. H. Hoess, W. F. DeGrado, *Science* **249**, 774 (1990).
30. M. G. Oakley and P. B. Dervan, *Science* **248**, 847 (1990).
31. T. G. Oas and P. S. Kim, *Nature* **336**, 42 (1988); J. P. Staley and P. S. Kim, *ibid.* **344**, 685 (1990).
32. CD spectra were obtained with an AVIV model 60HDS CD spectrometer at 25°C in a 5-mm cell. Samples contained 10 mM phosphate buffer (pH 7.0), 100 mM NaCl, and 4.6 μM GCN4-brl^{ss} and 5.0 μM GRE20 when present. Spectra in (A) and (B) were the average of multiple scans and were baseline-corrected with a spectrum of buffer alone, but were not smoothed.
33. Plasmid pUC9-Sc4251, containing the GRE sequence (*25*), has the 1.3-kb Eco RI–Bam HI fragment of plasmid YIp55-Sc4251 (*25*) cloned into the Eco RI–Bam HI site of pUC9 and was kindly provided by K. Struhl.
34. We thank A. Frankel for advice and discussions in all aspects of this work, E. O'Shea for preliminary experiments and discussions, R. Rutkowski for expert peptide synthesis, and S. Stradley and L. Gierasch for performing quantitative amino acid analysis. Supported by National Research Service Award GM13665 from the National Institutes of Health (R.V.T.), a postdoctoral fellowship from the Massachusetts Division of the American Cancer Society (C.J.M.), and by grants from the National Institutes of Health (GM44162) and the Lucille P. Markey Charitable Trust (P.S.K.).

1 May 1990; accepted 22 June 1990

Evidence of Changes in Protease Sensitivity and Subunit Exchange Rate on DNA Binding by C/EBP

JON D. SHUMAN, CHARLES R. VINSON, STEVEN L. MCKNIGHT

The transcription factor C/EBP uses a bipartite structural motif to bind DNA. Two protein chains dimerize through a set of amphipathic α helices termed the leucine zipper. Highly basic polypeptide regions emerge from the zipper to form a linked set of DNA contact surfaces. In the recently proposed a "scissors grip" model, the paired set of basic regions begin DNA contact at a central point and track in opposite directions along the major groove, forming a molecular clamp around DNA. This model predicts that C/EBP must undertake significant changes in protein conformation as it binds and releases DNA. The basic region of ligand-free C/EBP is highly sensitive to protease digestion. Pronounced resistance to proteolysis occurred when C/EBP associated with its specific DNA substrate. Sequencing of discrete proteolytic fragments showed that prominent sites for proteolysis occur at two junction points predicted by the "scissors grip" model. One junction corresponds to the cleft where the basic regions emerge from the leucine zipper. The other corresponds to a localized nonhelical segment that has been hypothesized to contain an N-cap and facilitate the sharp angulation necessary for the basic region to track continuously in the major groove of DNA.

THE TRANSCRIPTION FACTOR C/EBP regulates gene expression in a variety of tissues, including liver, adipose, lung, and intestine. The protein binds DNA through a bipartite structural motif consisting of a dimer-forming region immediately preceded by a polypeptide region rich in basic amino acids. Leucine residues occur in a heptad array along the dimer interface. Anticipating that the leucine residues would provide attractive, intersubunit interactions, we termed the dimer-forming region the leucine zipper (*1*). Biophysical studies have documented the α-helical nature of the leucine zipper and have shown that helices intertwine around one another in a parallel orientation (*2*). Considerable evidence has confirmed the role of the leucine zipper in dimerization of both identical and nonidentical protein subunits (*3*).

A variety of observations on transcription factors of this class have indicated that direct contact with DNA is mediated by the basic region. For example, a chimeric protein containing the basic region of C/EBP linked to the leucine zipper of GCN4 binds DNA with the specificity of C/EBP (*4*).

Proteins that use the contiguous basic region–leucine zipper arrangement (bZIP proteins) exhibit an invariant, six–amino acid spacing between the two components. Noting this fixed spatial register, as well as an absence of Pro and Gly residues, Vinson and colleagues (*5*) predicted that the basic region, like the zipper, would adopt an α-helical conformation. DNA-bound protein was hypothesized to form a Y-shaped molecule, the stem and arms corresponding, respectively, to paired zippers and bifurcating basic regions. This arrangement allowed the two basic regions to penetrate the major groove of DNA from a common point (the cleft of the Y), then track in opposite directions along each half of a dyad-symmetric binding site. Finally, this modeling predicted that α-helical structure would be locally disrupted within the basic region, facilitating a sharp bend necessary to allow continuous tracking of each basic region around the DNA on the side opposite to initial entry.

This model for bZIP proteins has been compared to the "scissors grip" hold that a wrestler uses to grasp the torso of an opponent. By wrapping around the DNA molecule on the side opposite of initial entry, the two subunits of a bZIP protein form a molecular clamp. If correct, this model demands that the protein undertake significant conformational changes as it binds and releases DNA. It further predicts that subunit exchange, which occurs rapidly in the absence of DNA, should be slowed dramatically upon DNA binding.

We examined the susceptibility of C/EBP to trypsin cleavage in the presence and absence of its DNA substrate. Trypsin, which cleaves the peptide bond carboxyl terminal to Arg and Lys residues, is a sensitive probe of the folded state (*6*). Moreover, C/EBP contains eight potential sites for trypsin cleavage in its basic region, six in its leucine zipper, and two in the short segment that links the basic region to the zipper (Fig. 1A).

Purified C/EBP (*7*) was exposed for 1-min intervals to varying amounts of trypsin. Digestion products were separated by electrophoresis on an SDS-polyacrylamide gel, transferred to nitrocellulose, and detected by immunoblotting with an antibody (α-C) specific to the carboxyl terminus of C/EBP (*8*). This strategy (*9*) provided a fixed labeling site on C/EBP, thus allowing a reasonably accurate identification of the sites of trypsin cleavage.

The patterns of trypsin cleavage of C/EBP alone, or of protein samples that had been mixed with either nonspecific or specific

Howard Hughes Research Laboratories, Department of Embryology, Carnegie Institution of Washington, Baltimore, MD 21210.

substrate DNA, are shown in Fig. 1B (*10*). Similar patterns of proteolysis were observed upon digestion of free protein or protein that had been mixed with nonspecific DNA. The earliest cleavage events occurred upon addition of between 30 and 100 ng of trypsin and appeared to be located in the amino-terminal half of the basic region. Addition of 100 to 300 ng of trypsin led to extensive digestion throughout the basic region. Finally, addition of 1 μg of trypsin removed the entire basic region, leaving a small amount of a 5-kD fragment consisting of the leucine zipper and carboxyl-terminal epitope.

A different pattern of proteolysis was observed when C/EBP was mixed with its specific DNA substrate. Little or no cleavage occurred until addition of 1 μg of trypsin. The limited amount of proteolysis detectable under these conditions appeared to occur at the junction between the basic region and the leucine zipper.

To accurately identify the sites of proteolysis observed under the conditions outlined in Fig. 1B, we separated digestion products by reversed-phase high-performance liquid chromatography (HPLC). By examining chromatograms corresponding to each of the digestion series shown in Fig. 1B, we identified and isolated individual proteolyzed species (*11*). For example, the trypsin digestion series conducted on the C/EBP sample that had been mixed with specific DNA substrate generated a discernible proteolytic product only after exposure to 1 μg of trypsin. This product, which was observed to elute from the HPLC column with a retention time roughly 1.5 min longer than the undigested sample, was recovered and sequenced in its entirety. It started at Asn^{307} of C/EBP and ended with the same carboxyl terminus as was present on the undigested sample (*12*).

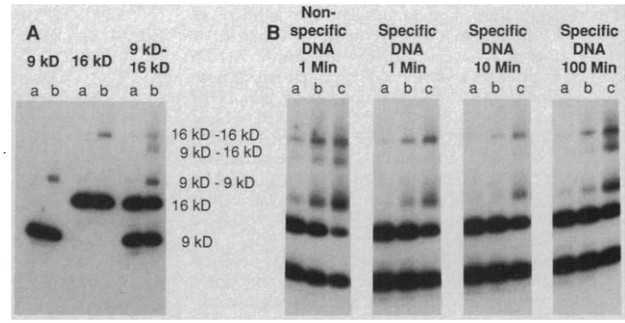

Fig. 2. Measurements of C/EBP subunit exchange rate in the presence and absence of DNA. (**A**) Patterns of glutaraldehyde-mediated cross-linking of a 9-kD fragment of C/EBP, a 16-kD fragment, and a sample of the two fragments that had been mixed at room temperature for 1 min prior to cross-linking. Lanes labeled (a) were loaded with protein that was not exposed to glutaraldehyde, and lanes labeled (b) were loaded with protein that was exposed to 0.005% glutaraldehyde for 1 min at room temperature. Deduced monomeric and heteromeric forms are indicated on the right. (**B**) Patterns of glutaraldehyde-mediated cross-linking observed for 9-kD and 16-kD protein fragments that had been exposed to either nonspecific or specific DNA before subunit mixing. Proteins that had been exposed to specific DNA were mixed for either 1, 10, or 100 min prior to cross-linking. Gel lanes were loaded with protein samples that had been exposed to either (a) 0.01%, (b) 0.02%, or (c) 0.04% glutaraldehyde.

The Asn^{307} residue directly follows an Arg residue and occurs in the polypeptide region that links the basic region of C/EBP to its leucine zipper (Fig. 1A). Sequencing of additional proteolytic fragments confirmed the interpretation that the earliest trypsin cleavages occur in the amino-terminal half of the basic region. Two prominent sites of cleavage generated peptides bearing amino termini at Glu^{290} (which directly follows two Arg residues) and Asn^{292} (which directly follows an Arg residue).

The DNA binding domain of C/EBP becomes substantially trypsin resistant when complexed with its ligand. As pointed out in the classical studies of Anfinsen and colleagues (*6*), ligand-induced protection against proteolysis can be explained in two ways. The bound ligand might directly mask access of the protease for a cleavage site. Alternatively, ligand binding can "rigidify" a protein, causing it to exist for a larger fraction of time in its appropriately folded state.

Given the extensive molecular contact expected to occur between the basic region of C/EBP and its DNA substrate (*5*), some degree of protease protection almost certainly results from direct steric masking by the DNA ligand. However, it is unlikely that DNA is simultaneously able to mask all eight potential trypsin cleavage sites in the C/EBP basic region. Indeed, the two cleavage sites that trypsin recognizes most readily (Arg^{289}-Glu^{290} and Arg^{291}-Asn^{292}) are predicted to be exposed on the surface of the protein-DNA complex (*5*). We therefore believe that substrate binding causes the basic region of C/EBP to adopt an ordered, trypsin-resistant conformation. This interpretation is consistent with the results of circular dichroism spectroscopic studies that show that the α-helical content of the C/EBP DNA binding domain doubles upon its association with specific DNA substrate (*13*).

According to the scissors grip model, the paired set of basic regions adopt an α-helical structure and track in the major groove around the side of DNA opposite to the leucine zipper. Continuous tracking of α helices demands a pronounced angulation at some point within the basic region. An N-cap has been hypothesized to form at Asn^{292} of C/EBP, allowing a localized disruption of α-helical structure within the basic region. This Asn residue is present midway through the basic region of all bZIP proteins described thus far. Perhaps the most convincing evidence supporting the scissors grip model derives from the locations of trypsin cleavage within the basic region. Recall that

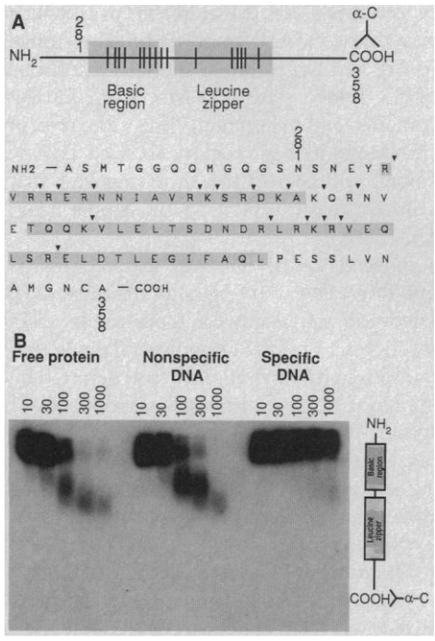

Fig. 1. Trypsin cleavage patterns of C/EBP in the absence and presence of DNA substrate. (**A**) Schematic diagram of the C/EBP fragment that was used in protease digestion experiments along with its amino acid sequence. Numbers above Asn residue (281) and below Ala residue (358) refer to amino acid positions within intact C/EBP (7). Potential trypsin cleavage sites are designated by vertical lines on the top diagram and inverted carets on the carboxyl-terminal side of each Arg and Lys in the bottom diagram. Intact and proteolytically digested products were detected by immunoblotting with an antibody specific to the carboxyl terminus of C/EBP (α-C). Limits of the basic region and leucine zipper are designated by stippling in both diagrams. 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. (**B**) Patterns of trypsin digestion of C/EBP in the absence of DNA, and in the presence of nonspecific and specific DNA substrate. Numbers above each gel lane refer to the amount of trypsin in nanograms that was added to each digestion reaction (*8*).

the two predominant sites of trypsin cleavage on C/EBP were located at Arg^{289}-Glu^{290} and Arg^{291}-Asn^{292}. These two peptide bonds co-localize with the hypothetical N-cap and localized nonhelical segment of the basic region (*5*).

Despite containing five potential trypsin cleavage sites, the leucine zipper of C/EBP was considerably resistant to proteolysis even in the absence of DNA substrate. We assume that the leucine zipper is folded in an α-helical conformation and that its well-ordered structure blocks proteolytic attack (*14*). Although the leucine zipper of ligand-free C/EBP was more resistant to proteolysis than the basic region, addition of large amounts of trypsin resulted in its near-complete digestion (Fig. 1B). When complexed with specific DNA substrate, the entire DNA binding domain, including the leucine zipper, became protease resistant.

C/EBP subunits exchange readily in the absence of DNA (*15*), perhaps offering transient access to proteolytic attack. If so, DNA binding might, by retarding the rate of subunit interchange, indirectly confer protease resistance to the zipper. In order to compare subunit exchange rates in the presence and absence of DNA, two different fragments of C/EBP were prepared. One fragment consisted of amino acid residues 281 to 358 of C/EBP and migrated on SDS-polyacrylamide gels with an apparent molecular weight of 9 kD. The other fragment consisted of residues 214 to 358 and exhibited an apparent molecular weight of 16 kD. As shown in Fig. 2A, brief exposure of the smaller fragment to low concentrations of glutaraldehyde generated an 18-kD form. Likewise, glutaraldehyde treatment of the larger protein generated as 32-kD form. If the two fragments were mixed for 1 min at room temperature before exposure to glutaraldehyde, an additional cross-linked form was generated. As argued in an earlier study (*15*), we interpret the additional form to represent a heteromeric pair consisting of one short polypeptide chain and one long chain.

Subunit exchange experiments were conducted with pure protein samples as well as samples that had been preincubated with either specific or nonspecific DNA substrate (*16*). As shown in Fig. 2B, addition of the specific DNA substrate markedly retarded the rate of subunit interchange. The presumed heteromeric species began to be detected only after a 100-min mixing interval. Conversely, in the absence of DNA, or in the presence of nonspecific substrate, subunits were observed to interchange within 1 min.

The results of these experiments indicate that subunit exchange is slowed substantially when C/EBP is bound to its specific DNA substrate. Detection of exchange only after a 100-min mixing interval does not, however, mean that C/EBP stays permanently associated with DNA. When DNA-bound C/EBP is challenged with excess substrate, the protein dissociates within 5 min (*17*). We instead believe that under the assay conditions used in this experiment, the rate of association between C/EBP and its specific substrate is considerably more rapid than the rate of subunit interchange. Thus, when C/EBP dissociates from DNA, it is more likely to reassociate with DNA than undergo subunit interchange.

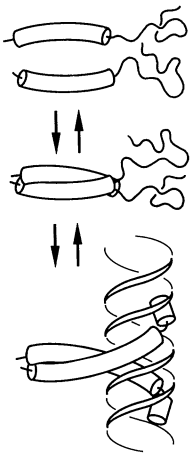

Fig. 3. Two-step binding reaction between C/EBP and DNA substrate. In the absence of DNA, polypeptide subunits of C/EBP exist in equilibrium between the monomeric and dimeric state. Subunit association is thought to be mediated by the leucine zipper. Coiled tubes (left) correspond to α helices of paired leucine zippers. DNA binding causes C/EBP subunits to become trypsin-resistant and is hypothesized to lead to changes in the conformation of the basic region. Prior to DNA binding, basic regions are trypsin-sensitive and presumed to occur for a significant fraction of time as random coils. When bound to DNA the basic regions become trypsin-resistant and are presumed to occur as well-ordered α helices. Subunit exchange is substantially retarded when protein is bound to specific substrate.

The experiments outlined in this study provide evidence of two macromolecular interactions that may occur in a stepwise manner to facilitate DNA binding by C/EBP (Fig. 3). In the first step, protein subunits coalesce through the leucine zipper. In the absence of specific DNA substrate, this step is readily reversible. The second step that we tentatively identify involves the binding of C/EBP to its DNA substrate. Two properties of C/EBP appear to change upon completion of this second step. First, the basic region acquires resistance to cleavage by trypsin. Second, the exchange rate of protein subunits is retarded.

We have speculated that the substrate-dependent acquisition of protease resistance in the basic region of C/EBP reflects, at least in part, a transition in protein conformation from random coil to α helix (Fig. 3). A related example of "induced fit" between protein and nucleic acid has been reported by Warrant and Kim (*18*), who found that protamines undertake a random coil to α helix transition upon exposure to crystalline tRNA. A concerted network of attractive charge interactions between basic amino acids of C/EBP and the negatively charged DNA substrate may induce the formation of α-helical structure within the basic region. Since the acquisition of trypsin resistance is DNA sequence dependent, the conformational changes that we hypothesize may relate to the manner in which C/EBP is able to discriminate between different sites on DNA.

REFERENCES AND NOTES

1. W. H. Landschulz, P. F. Johnson, S. L. McKnight, *Science* **240**, 1759 (1988).
2. E. K. O'Shea, R. Rutkowski, P. S. Kim, *ibid.* **243**, 538 (1989); T. G. Oas, L. P. McIntosh, E. K. O'Shea, F. W. Dahlquist, P. S. Kim, *Biochemistry* **29**, 2821 (1990).
3. T. Abel and T. Maniatis, *Nature* **341**, 24 (1989); T. Kouzarides and E. Ziff, *Cancer Cells* **1**, 71 (1989); N. Jones, *Cell* **61**, 9 (1990).
4. P. Agre, P. F. Johnson, S. L. McKnight, *Science* **246**, 922 (1989).
5. C. R. Vinson, P. B. Sigler, S. L. McKnight, *ibid.*, p. 911.
6. H. Taniuchi and C. B. Anfinsen, *J. Biol. Chem.* **243**, 4778 (1968); H. Taniuchi, L. Moravek, C. B. Anfinsen, *ibid.* **244**, 4600 (1969).
7. Three polypeptide fragments containing the DNA binding domain of C/EBP were used in the experiments shown in Figs. 1 and 2. The smallest fragment consisted of 62 amino acids and corresponded to residues 281 to 340 of the native C/EBP sequence [W. H. Landschulz, P. F. Johnson, E. Y. Adashi, B. J. Graves, S. L. McKnight, *Genes Dev.* **2**, 786 (1988)]. The 62-residue polypeptide contained, at its amino terminus, two amino acids derived from the T7 expression vector (Gly Ser). A second fragment consisted of 92 amino acids and corresponded to residues 281 to 358 of the native sequence. The gene encoding this 92-residue fragment was altered by in vitro mutagenesis [M. J. Zoller and M. Smith, *Methods Enzymol.* **100**, 469 (1983)] to remove three potential trypsin cleavage sites close to the carboxyl terminus of C/EBP (Arg^{339} to Glu, Arg^{343} to Ala, and Lys^{352} to Asn). This second fragment also contained, at its amino terminus, 14 residues derived from the amino terminus of the gene 10 protein of bacteriophage T7. The largest fragment consisted of 149 amino acids and corresponded to residues 214 to 358 of the native C/EBP sequence. The amino terminus of this larger polypeptide also contained two residues derived from the T7 expression vector (Gly Ser). Proteins were synthesized in *Escherichia coli* with the phage T7 expression system [F. W. Studier and B. A. Moffatt, *J. Mol. Biol.* **189**, 113 (1986)]. In each case a 3-liter culture was grown in a Lab-Line/SMS fermenter to an optical density at 600 nm of 10. Synthesis of C/EBP was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to 2 mM. Cells were harvested 2 hours later, recovered by centrifugation, and resuspended in 80 ml of buffer A [10 mM tris-HCl, pH 7.9, 100 mM KCl, 0.5 mM EDTA, 1 mM benzamidine, 1 mM dithiothreitol (DTT), and 0.2 mM phenylmethylsulfonyl fluoride (PMSF)] supplemented with 1 mM $MgCl_2$, 2 mM $CaCl_2$, lysozyme (1 μg/ml), and deoxyribonuclease I (10 μg/ml). After 20 min of shaking at 4°C, the cell suspension was freeze-thawed twice and sonicated to insure complete lysis. Ammonium sulfate was added to 25% (weight per volume), and the lysate was centrifuged at 7000 rpm for 60 min in a Sorvall GS3 rotor. Supernatant was dialyzed against buffer A, bound to a 300-ml column of S-Sepharose Fast Flow (Pharmacia), and eluted with a linear salt gradient from 0.1 to 0.8 M KCl in buffer A. The C/EBP-containing fractions were pooled, dialyzed against 0.5× buffer A, brought to 1.2 M ammonium sulfate, bound to an

80-ml column of phenyl Sepharose (Pharmacia), and eluted with a linear salt gradient from 1.2 to 0 M ammonium sulfate in buffer A. Fractions containing C/EBP were pooled, dialyzed against 0.5× buffer A, bound to an 80-ml column of calf thymus DNA Sepharose [P. F. Johnson, W. H. Landschulz, B. J. Graves, S. L. McKnight, *Genes Dev.* **1**, 133 (1987)], and eluted with a linear salt gradient from 0.05 to 0.7 M KCl in buffer A. All three forms of C/EBP were judged to be greater than 99% pure as determined by silver staining of SDS-polyacrylamide gels and reversed-phase HPLC.

8. Free C/EBP and protein samples that had been mixed with DNA were exposed to trypsin in 20-μl reactions containing 25 mM tris-HCl, pH 7.9, 50 mM KCl, 0.5 mM DTT, and 0.5 mM EDTA. Each reaction contained 1.2 μg of the 92-residue form of C/EBP, and, where indicated, a twofold molar excess of DNA (*10*). Trypsin (Boehringer) was freshly diluted, added in amounts ranging from 10 ng to 1 μg, and stopped by the addition of SDS gel sample buffer. The products of trypsin digestion were separated by electrophoresis on an 18% polyacrylamide gel [U. K. Laemmli, *Nature* **227**, 680 (1970)], transferred to nitrocellulose, and detected by sequential incubation of the filter with an antibody specific to the 15 residues at the carboxyl terminus of C/EBP [E. H. Birkenmeier *et al.*, *Genes Dev.* **3**, 1146 (1989)], and ^{125}I-labeled Protein A (Amersham).
9. P. Matsudaira, R. Jakes, L. Cameron, E. Atherton, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6788 (1985).
10. Ten micromoles of internally symmetric, 20-residue oligonucleotides containing either a high-affinity C/EBP binding site (5′-TGCAGATTGCGCAATC-TGCA-3′) or unrelated sequence (5′-TGCAGAGA-CTAGTCTCTGCA-3′) were synthesized and purified by two passages over a reversed-phase HPLC C18 column. Oligonucleotides were lyophilized following the second HPLC pass, resuspended in and dialyzed against 1 mM tris-HCl, pH 8.0, 100 mM KCl, and annealed by heating to 95°C and allowing the reaction to cool slowly to 45°C. The concentration of double-stranded product was determined spectrophotometrically and adjusted to 5 mg/ml by addition of buffer.
11. Trypsin cleavage reactions were performed as described (*8*) with the 62-residue form of C/EBP (*7*). Reactions were terminated by the addition of PMSF to 20 mM and injected onto a Vydac 2.1 mm by 250 mm C18 reverse-phase column. Proteolytic fragments were eluted with a linear gradient from 0 to 50% acetonitrile in 0.1% trifluoroacetic acid. The undigested 62-residue C/EBP eluted with a retention time of 37 min. Proteolytic fragments chosen for amino acid sequencing eluted at retention times of 38.5, 17.2, and 17.6 min.
12. Individual peaks collected following HPLC separation were lyophilized to a volume of 100 μl and sequenced with an Applied Biosystems Model 477A pulsed liquid-phase protein sequencer.
13. K. O'Neil, J. Shuman, W. F. DeGrado, unpublished observations.
14. E. Mihalyi and W. F. Harrington, *Biochim. Biophys. Acta* **44**, 191 (1960); J. A. Rupley and H. A. Scheraga, *ibid.*, p. 191.
15. W. H. Landschulz, P. F. Johnson, S. L. McKnight, *Science* **243**, 1681 (1989).
16. Glutaraldehyde-mediated cross-linking was performed as described (*15*). Reactions contained 700 nmol of C/EBP, composed of either 9-kD protein, 16-kD protein, or an equimolar mix of the two. Oligonucleotides were added at a twofold molar excess over protein and incubated at room temperature for 10 min prior to subunit mixing. Cross-linking was terminated by the addition of SDS gel sample buffer. Unreacted and cross-linked protein species were detected by immunoblotting after electrophoretic separation on SDS-polyacrylamide gels (*8*).
17. C. Vinson, unpublished observations.
18. R. W. Warrant and S.-H. Kim, *Nature* **271**, 130 (1978).

19. We thank K. Kirkegaard for suggesting the experimental strategy used in Fig. 1, B. Byers for extensive help in the synthesis and purification of oligonucleotides and in the resolution, recovery, and sequencing of proteolyzed products of C/EBP, P. Sigler for interest and critical input, B. Kingsbury for technical help, C. Norman for clerical help, and our colleagues at the Carnegie Embryology Department for advice and encouragement. J.D.S. and C.R.V. were supported by fellowship stipends from the Leukemia Society and American Cancer Society, respectively. This work was otherwise funded by the Howard Hughes Medical Research Institute and the Carnegie Institution of Washington.

2 May 1990; accepted 20 June 1990

Design of DNA-Binding Peptides Based on the Leucine Zipper Motif

KARYN T. O'NEIL, RONALD H. HOESS, WILLIAM F. DEGRADO

A class of transcriptional regulator proteins bind to DNA at dyad-symmetric sites through a motif consisting of (i) a "leucine zipper" sequence that associates into noncovalent, parallel, α-helical dimers and (ii) a covalently connected basic region necessary for binding DNA. The basic regions are predicted to be disordered in the absence of DNA and to form α helices when bound to DNA. These helices bind in the major groove forming multiple hydrogen-bonded and van der Waals contacts with the nucleotide bases. To test this model, two peptides were designed that were identical to natural leucine zipper proteins only at positions hypothesized to be critical for dimerization and DNA recognition. The peptides form dimers that bind specifically to DNA with their basic regions in α-helical conformations.

THE LEUCINE ZIPPER CLASS OF transcriptional regulators (*1*) are characterized by a pattern of Leu residues repeating every seventh amino acid (*2*). This motif mediates protein dimerization (*3, 4*) through the formation of parallel α-helical dimers as in the two-stranded coiled coils of fibrous proteins (*3*). At the NH_2-terminus of the leucine zipper begins a highly basic region 20 to 30 residues in length that is required for specific binding to DNA (*5, 6*). To explore how this class of proteins bind specifically to DNA we have applied a "minimalist" approach to protein design (*7*), which seeks to find the simplest structure consistent with a given function.

In order to create models for this class of proteins, we first predicted the secondary structure of their basic regions by searching for patterns of conserved and variable residues in their sequences (Fig. 1A). Side chains important for protein folding or ligand binding often lie along one face of a secondary structure and are highly conserved (*8*), giving a pattern of conserved residues with a repeat characteristic of the corresponding secondary structure, approximately 2 for β sheets or 3 to 4 for α helices. In Fig. 1B a plot is shown of the degree of variability along the chain for the aligned sequences of basic regions that bind to one of two similar dyad-symmetric sites, TGAC/GTCA (TRE) or TGACGTCA (CRE). We assume that each of these proteins contains a similar set of side chains responsible for their binding to the conserved TGA sequence. A periodic distribution is observed with a repeat of 3.6, matching that of the α helix.

Among the most conserved residues in the sequence were the neutral amino acids, Asn^{-18}, Ala^{-15}, Ala^{-14}, and Cys^{-11} (numbering is relative to the first Leu repeat, immediately COOH-terminal to the basic region). Each of these residues is absolutely conserved with the exception of Ser^{-11} for Cys^{-11} (an OH for SH substitution), and at least one member of this conserved quartet differs in other leucine zipper proteins with different DNA specificities. The conserved quartet is also bounded on all sides by positively charged Lys and Arg side chains (Fig. 1C). In specific complexes, the conserved quartet might directly interact with the bases. The short side chains of the conserved quartet would provide relatively little conformational freedom, possibly important for minimizing the unfavorable conformational entropy of binding and decreasing the likelihood of recognizing other DNA sequences. The positively charged residues flanking the conserved quartet might stabilize the complex through electrostatic interactions with the phosphodiester backbone. Interestingly, two acidic Glu residues often occur on the otherwise variable face of the helix and might electrostatically stabilize the helical conformation of the cationic basic region (*9*), and, relative to the positively

Central Research and Development Department, E. I. du Pont de Nemours & Co., P.O. Box 80328, Wilmington, DE 19880–0328.