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.

charged Arg and Lys residues, create a dipole with its positive pole pointing toward DNA (*10*).

A model describing the interaction of the basic region with DNA should be consistent with the following experimental observations: (i) the DNA site is approximately twofold symmetric; (ii) ethylation interference experiments indicate that approximately ten phosphates are contacted per strand (*6*); and (iii) methylation interference experiments indicate that binding occurs in the major groove (*6, 11, 12*). We further assume that the dimer is approximately twofold symmetric and that it binds with its twofold axis (defined by the coiled coil of the leucine zipper) coincident with that of the inverted repeat, positioning the COOH-terminus of the basic region near the DNA sequence dyad. We also assume that the conserved quartet directly contacts the DNA bases. We devised a simple model (Fig. 2) that satisfies these constraints in which the basic region forms α helices continuous with the leucine zipper helices. In this model, the leucine zipper sequence forms an α-helical coiled coil with an interhelical crossover angle of about 20°. To avoid diverging, the helices in the coiled-coil wrap around one another forming a supercoil. If the supercoiling stops at the junction of the basic region and the leucine zipper, the helices diverge producing a fork that can bind with the conserved quartet contacting the major groove of DNA.

In this model the basic region helix is not stabilized by tertiary interactions with the remainder of the protein, in contrast to the recognition sequences of other site-specific DNA-binding proteins of known structure (*13*). Our model does not consider other portions of the protein that lie outside of basic region–leucine zipper motif and might form tertiary interactions. However, it is likely that such interactions with the remainder of the protein are thermodynamically unimportant for binding DNA because: (i) footprints, specificities, and affinities for intact zipper proteins and fragments encompassing the DNA-binding motif are nearly identical (*14*); and (ii) the faces of helices involved in tertiary interactions generally show a reasonable degree of hydrophobicity and sequence conservation (*8*), which is not the case for the basic region if the conserved quartet contacts DNA. Thus, we would predict that the basic region would form a helix only in the presence of the DNA, and refer to our model as the induced helical fork because the active conformation (resembling a tuning fork) is hypothesized to be induced in the presence of DNA.

Subsequent to the completion of our model, a "scissors grip" model was independently reported for the complex of the basic region–leucine zipper (*11*). Although the reasoning used to create this model was different from that described in Fig. 2, the overall structural features of the two models are fundamentally similar. One major difference between the two is that the scissors grip does not predict which residues specifically interact with DNA. Also, the scissors grip model hypothesizes that Asn^{-18} simultaneously hydrogen bonds to the phosphodiester backbone of DNA and the peptide backbone of the basic region helix, inducing a kink in the α helix. While the specific

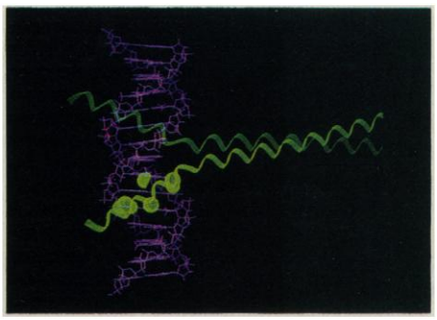
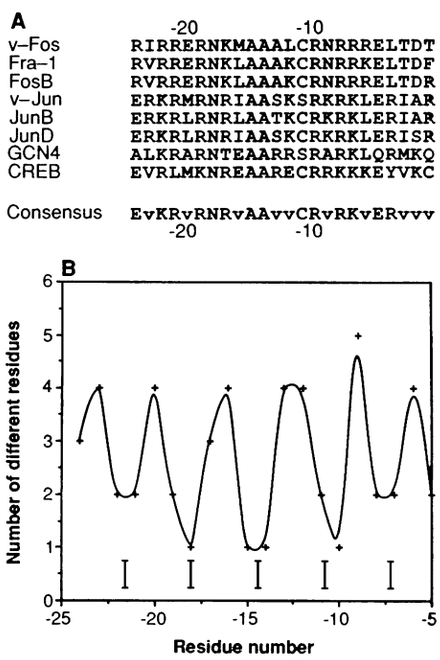


Fig. 2. Computer model of the induced helical fork. The DNA is shown in purple, the peptide in green, and van der Waals surfaces highlight the conserved quartet for one-half of the dimer. A B-form DNA helix (*25*) was generated with the sequence (CCTATGACTCATCCA and its complement). The backbone of the basic region was extracted from a curved helix (residues 79 to 102) of cytochrome C′ (*26*). The side chains of the basic region from GCN4 were built onto this backbone with INSIGHT (Biosym Technologies, Inc.) in conformations typically found in proteins (*27*). The basic region helix was oriented in the major groove to form: (i) a bidentate hydrogen bond (*13*) between Asn^{-18} and the adenine complementary to the first thymine in the half-site, $T_1G_2A_3$; and (ii) a bifurcated hydrogen bond between the Ser^{-11} hydroxyl proton and O6 and N7 of G_2. This orientation places steric restrictions on the side chains at −14 and −15 (both are Ala). The coiled-coil portion (*28*) could be readily attached to the basic region with minimal perturbation of the helical parameters and in the same registry as observed in natural leucine zipper proteins.

```
BR-CC   A L K R A R N T E A A R R S R A R K L Q R M K Q L E D K
CC
MBR-CC  E A R R A R N R E A A A R S R A R R A E K L K A L E E K
GCN4    A L K R A R N T E A A R R S R A R K L Q R M K Q L E D K
Jun-D   E R K R L R N R I A A S K C R K R K L E R I S R L E E K
c-Fos   R I R R E R N K M A A A K C R N R R R E L T D T L Q A E

BR-CC   V K E L E E K L K A L E E K L K A L E E K L K A L G
CC      V K E L E E K L K A L E E K L K A L E E K L K A L G
MBR-CC  L K A L E E K L K A L E E K L K A L E E K L K A L G
GCN4    V E E L L S K N Y H L E N E V A R L K K L V G E R
Jun-D   V K T L K S Q N T E L A S T A S L L R E Q V A Q L K
c-Fos   T D Q L E D E K S A L Q T E I A N L L K E K E K L E
```

Fig. 3. Sequences of idealized peptides and the corresponding sequences from GCN4, c-Fos, and Jun-D. The peptides were synthesized with a Milligen (Burlington, Massachusetts) 9050 automated peptide synthesizer by using Fmoc amino acids, Fmoc-PAL resin, and standard procedures described by Milligen. Purification was accomplished by reversed-phase high-performance liquid chromatography with a semipreparative Vydac C18 column, flow rate 3 ml/min with the following gradients (percentages refer to acetonitrile containing 0.1% trifluoroacetic acid): CC, 36 to 45% over 26 min; BR-CC, 32 to 43% over 20 min; and MBR-CC, 36 to 41% over 20 min. The peptides were homogeneous by criteria of amino acid analysis and fast atom bombardment mass spectrometry.



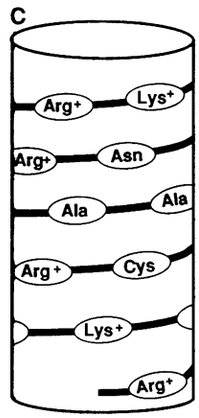


Fig. 1. (**A**) Amino acid sequences (*24*) from the basic regions of eight proteins that bind to either the TRE or CRE site. Sequences are numbered relative to the first leucine in the zipper portion of each protein. The bottom row illustrates a consensus sequence. Amino acids conserved in at least one-half the proteins are identified, whereas more variable positions are denoted with the letter v. (**B**) Variability at each position in the basic region. The number of different amino acids allowed at a given position is plotted as a function of position along the chain. The intervals shown at the bottom mark the repeat for an α helix. (**C**) Graphical representation of the concensus sequence in an α-helical conformation showing the conserved face of the helix.

interactions formed by Asn^{-18} and other members of the conserved quartet can best be determined by NMR spectroscopy or x-ray crystallography, we have used the techniques of protein design and circular dichroism (CD) spectroscopy to provide strong evidence for the global structural features of both these models.

The induced helical fork model suggests that the role of the leucine zipper in DNA binding is simply to position the basic regions to interact with DNA, and possibly also to initiate helix formation in the basic region. Also, the model identifies residues in the basic region that should be important for helix stability but not DNA recognition. These residues might all be replaced with a small, neutral, helix-favoring amino acid. Therefore we designed a hybrid containing the basic region of GCN4 attached to an idealized coiled coil (Fig. 3). Hodges and co-workers have designed coiled coils based on the heptapeptide repeat (Leu-Glu-Gly-Lys-Leu-Glu-Ala) (*15*). We designed a similar repeat, Leu-Glu-Glu-Lys-Leu-Lys-Ala, which included Leu at analogous positions to hydrophobically stabilize the dimer, Lys and Glu to form stabilizing electrostatic interactions, and Ala to stabilize helix formation. Three repeats of the heptad provided the core of the coiled-coil peptide. The basic region plus the first heptad repeat of the leucine zipper of GCN4 were then added to the NH_2-terminus of the coiled coil giving the BR-CC (basic region–coiled coil) peptide. We chose the shortest basic region consistent with specific binding based on our model. As a control, another peptide, CC, lacking the basic region was also synthesized (Fig. 3).

A second peptide, MBR-CC (minimal BR-CC), was synthesized in which both the basic region and the leucine zipper were idealized. The conserved quartet was maintained, and Arg was included at positions predicted to be important for electrostatic interactions. Three Glu residues were included on the solvent exposed face of the helix, and all other positions were Ala. Four and one-half repeats of the idealized heptad formed the leucine zipper portion of MBR-CC. The degree of exact amino acid sequence homology between MBR-CC and the corresponding regions of GCN4, c-Fos, and Jun-D (Fig. 3) are 43, 31, and 31%, respectively. Of these conserved residues, only 11 (21%) are common to all four sequences.

The helical content of the peptides were estimated by CD. The spectra are typical of helical proteins with minima at 222 and 208 nm and a maximum at 190 nm. Quantitative analyses of the spectra by two methods (*16, 17*) indicated that CC and BR-CC were approximately 82 and 65% α-helical, respectively (Table 1), consistent with a helical zipper region and a largely unordered basic region. Analysis of the MBR-CC spectrum gave a slightly higher α helix content than BR-CC (75%).

We determined the conformational stability of the helical peptides by examining the variation in ellipticity at 222 nm (θ_{222}, a linear measure of helical content) with respect to the concentration of the guanidine ·

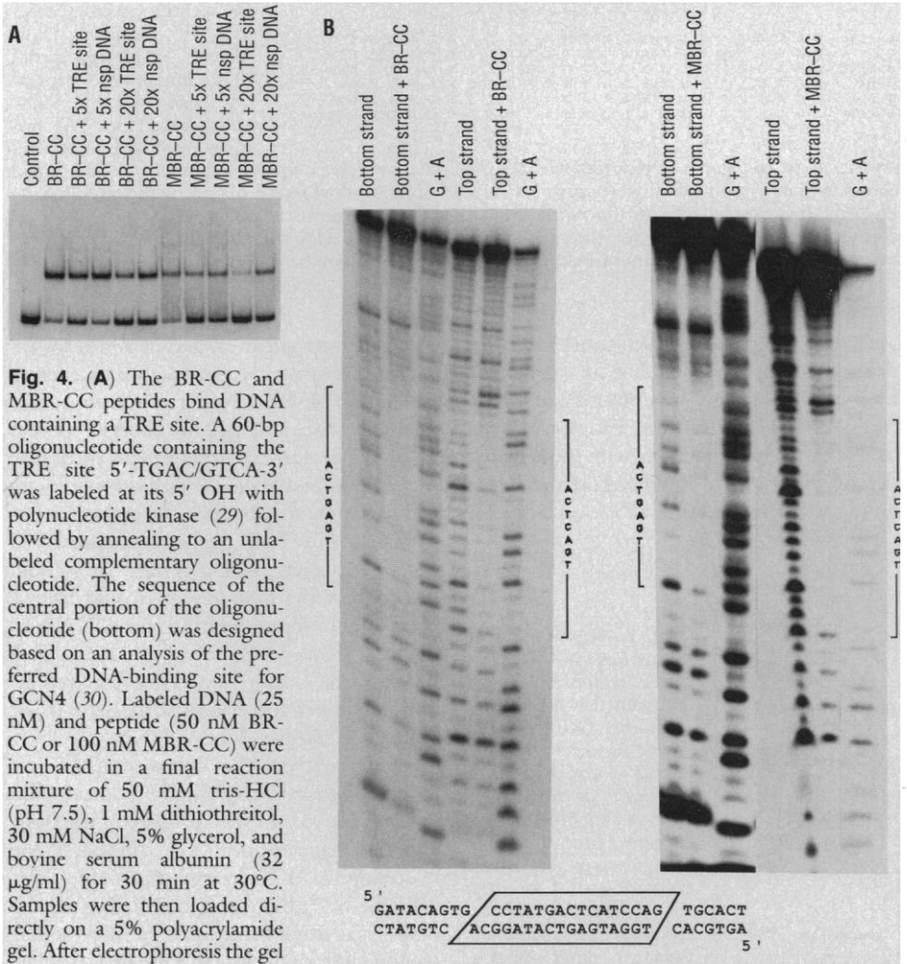


Fig. 4. (**A**) The BR-CC and MBR-CC peptides bind DNA containing a TRE site. A 60-bp oligonucleotide containing the TRE site 5′-TGAC/GTCA-3′ was labeled at its 5′ OH with polynucleotide kinase (*29*) followed by annealing to an unlabeled complementary oligonucleotide. The sequence of the central portion of the oligonucleotide (bottom) was designed based on an analysis of the preferred DNA-binding site for GCN4 (*30*). Labeled DNA (25 nM) and peptide (50 nM BR-CC or 100 nM MBR-CC) were incubated in a final reaction mixture of 50 mM tris-HCl (pH 7.5), 1 mM dithiothreitol, 30 mM NaCl, 5% glycerol, and bovine serum albumin (32 μg/ml) for 30 min at 30°C. Samples were then loaded directly on a 5% polyacrylamide gel. After electrophoresis the gel was fixed in 5% acetic acid and autoradiographed. Where indicated an excess of unlabeled 26-bp TRE site-containing oligonucleotide or a 26-bp oligonucleotide containing a mutated TRE site in (nsp DNA) were used as competitors. (**B**) Deoxyribonuclease I (DNase I) footprint of the TRE site by BR-CC or MBR-CC. Following binding of BR-CC or MBR-CC to the TRE site containing oligonucleotide under the conditions described in (A) with the addition of 5 mM $MgCl_2$, DNase I was added to the reaction mixture and incubated an additional 5 min at 30°C. Reactions were quenched by the addition of 0.6 M sodium acetate, 75 mM EDTA, phenol extracted, and ethanol precipitated prior to electrophoresis on a 10% polyacrylamide-urea gel. The G+A marker was generated as described previously (*29*). The brackets show the extent of the DNAse I footprint on the top and bottom strands. The protected DNA sequences are shown enclosed in the parallelogram below.

Table 1. Estimates of α helicity of peptides with or without DNA determined by two CD methods (*16, 17*). Stock solutions of peptide were prepared in 10 mM MOPS buffer, pH 7.5; peptide concentrations were determined by amino acid analysis. DNA concentrations were determined spectrophotometrically for the single strands with an extinction coefficient at 260 nm (ε260) of 40 $(\mu g/ml)^{-1}$. Strands were annealed by heating to 80°C for 10 min and then slowly cooling to room temperature; TRE, TRE site; nsp, nonspecific DNA. Spectra of the peptide dimers (10 to 25 μM) with or without a single equivalent of DNA duplex (given in parenthesis) were measured with a Jasco J-500 Spectropolarimeter in a 0.01-cm pathlength cell.

Peptide	$-\theta_{222}$ (deg cm^2 dmol)	α-Helical content (%)	
		(*16*)	(*17*)
CC	28,300	82	82
BR-CC	22,000	64	66
MBR-CC	25,700	75	75
BR-CC (nsp)	25,100	73	73
BR-CC (TRE)	30,950	90	88
MBR-CC (TRE)	31,700	92	90
BR	9,100	26	33
BR (nsp)	20,100	58	61
BR (TRE)	33,600	98	95

HCl (GuHCl). BR-CC and CC showed cooperative transitions with midpoints near 4 M GuHCl; MBR-CC unfolded at a slightly higher denaturant concentration (5 M). The dissociation constants for dimerization of the peptides in the absence of denaturant were estimated by the linear extrapolation method (*18*) to be approximately 4 pM for CC and 5 pM for BR-CC and MBR-CC. Thus the peptides dimerize with high affinity and should be entirely dimeric under conditions used to measure binding to DNA.

BR-CC bound specifically to a 60-bp ^{32}P-labeled oligonucleotide containing the TRE site, as assessed by the electrophoretic mobility shift assay (Fig. 4A). When CC was substituted for BR-CC, no bands indicative of complex formation were observed. MBR-CC binds to DNA analogously to BR-CC. An unlabeled oligonucleotide bearing the TRE site effectively competes for binding to the labeled DNA-peptide complexes, whereas an oligonucleotide containing a mutation of the TRE site (TTACTAA) was a much poorer competitor.

Binding of BR-CC protects the 60-bp oligonucleotide from deoxyribonuclease I cleavage at the 7-bp TRE site (Fig. 4B). As for other DNA binding proteins (*19*), the protection also extends 4 to 5 bases on either end of the recognition sequence. An identical footprint was observed for the MBR-CC peptide (Fig. 4B).

The helicity of the basic regions of BR-CC and MBR-CC increase in the presence of DNA, as assessed by CD. The CD spectra (Fig. 5A) of a 26-bp oligonucleotide containing the TRE site (20 μM) exhibits bands at approximately 270 and 240 nm, typical of DNA in the B form (*20*), both in the presence and absence of a single equivalent of the BR-CC dimer. Thus the peptide probably does not induce a major conformational change in the DNA. In contrast, the ellipticity of the peptide increases significantly upon interaction with DNA (Fig. 5A), corresponding to an increase to 90% helix content for BR-CC and MBR-CC (Table 1). This increase in helix content can be assigned primarily to the basic region as CC is nearly completely α-helical in the absence or presence of DNA. The difference spectra obtained by subtracting the spectrum of CC from that for BR-CC, either in the presence or absence of DNA, are shown in Fig. 5B. In the absence of DNA the spectrum is consistent with about 25% helix formation; the corresponding value in the presence of specific DNA is 95 to 100%. In the presence of a single equivalent of nonspecific oligonucleotide the increase in ellipticity was about one-half that induced by the specific oligonucleotide. Adding more nonspecific DNA failed to increase the ellipticity of the peptide as would be expected if the lower induction were the result of partial binding. Thus the peptide binds to nonspecific DNA but in a conformation with a lower helical content than for the specific complex.

The above experiments are consistent with models in which the basic region forms a helix, which binds to the major groove of DNA. Interestingly, the basic region binds nonspecific DNA in a partially helical conformation, possibly allowing the protein to efficiently "search" the DNA for its recognition sequence. Once the specific sequence is recognized the basic region becomes completely helical to allow tight binding interactions. To the best of our knowledge, mechanisms involving such large-scale conformational changes in the specificity-determining portions of DNA-binding proteins are unprecedented, although it has been known for many years that highly flexible peptides can bind rapidly and with high specificity to their receptors (*21*). For instance, S-peptide, a proteolytic fragment of ribonuclease, binds at a nearly diffusion controlled rate to its complementary fragment, S-protein (*22*). In the absence of S-protein, S-peptide is predominantly nonhelical under physiological conditions. The peptide initially binds in a nonhelical conformation, and the helix subsequently forms on a slower time scale (*23*).

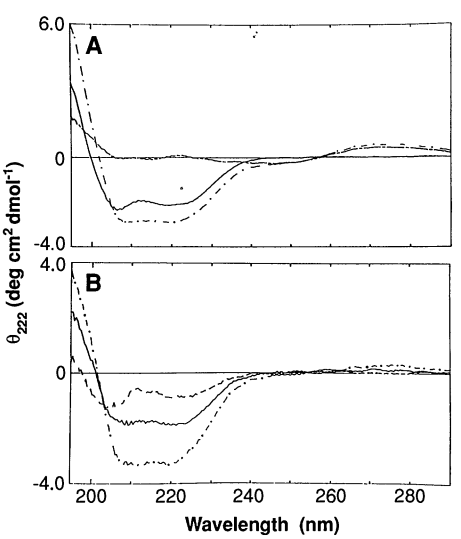


Fig. 5. (**A**) Induced CD of BR-CC by a 26-bp oligonucleotide containing the TRE site; free oligonucleotide (broken line), free peptide (solid line), and peptide plus DNA (dot-dash line). Spectra were measured using a 0.01-cm pathlength cell. Conditions: 10 mM MOPS, 0.15 M NaCl, pH 7.5, room temperature; peptide dimer concentration, 20 μM; and DNA concentration, 20 μM. Spectra were averaged five times and smoothed with the software provided by Jasco, Inc. (Easton, Maryland). (**B**) Difference spectra were obtained by subtracting the CC spectrum from the BR-CC spectrum at identical concentrations in the absence or presence of either a specific oligonucleotide or a random-sequence oligonucleotide. Data were collected and smoothed as described in (A).

Our results also show the potential of the minimalistic approach for the analysis of protein DNA interactions. We have used this technique to show that the leucine zipper–basic region motif is composed of two structural motifs: a coiled coil that can be replaced with a repetitive sequence; and a helical basic region, the proposed solvent-exposed face of which can sustain multiple sequence changes with minimal affect on the DNA specificity. We find our approach to be a useful complement to genetic approaches, which generally focus on less global changes to the amino acid sequence. The major advantage of mutagenesis is that it requires no structural knowledge of the protein of interest, and hence is useful for defining features likely to be critical for activity. However, once even limited mutagenic data are accumulated, hypotheses can be generated and rapidly tested with the minimal approach.

Note added in proof: We recently have shown that full-length wild-type GCN4 protein shows an increase in α-helical content when it binds specifically to DNA. The increase in θ_{222} observed when GCN4 is added to DNA is similar to that observed for BR-CC.

REFERENCES AND NOTES

1. P. F. Johnson and S. L. McKnight, *Annu. Rev. Biochem.* **58**, 799 (1989).
2. W. H. Landschultz, P. F. Johnson, S. L. McKnight, *Science* **240**, 1759 (1988).
3. E. K. O'Shea, R. Rutkowski, W. F. Stafford III, P. S. Kim, *ibid.* **245**, 646 (1989); E. K. O'Shea, R. Rutkowski, P. S. Kim, *ibid.* **243**, 538 (1989).
4. I. A. Hope and K. Struhl, *EMBO J.* **6**, 2781 (1987); ———, *Cell* **46**, 885 (1986); T. Kouzarides and E. Ziff, *Nature* **340**, 568 (1989); R. Turner and R. Tjian, *Science* **243**, 1689 (1989); P. Sassone-Corsi, L. J. Ransone, W. W. Lamph, I. M. Verma, *Nature* **336**, 692 (1988); T. Kouzarides and E. Ziff, *ibid.*, p. 646; L. J. Ransome, J. Visvader, P. Sassone-Corsi, I. M. Verma, *Genes Dev.* **3**, 770 (1989); R. Gentz, F. J. Rausher III, C. Abate, T. Curran, *Science* **243**, 1695 (1989); M. Schuermann *et al.*, *Cell* **56**, 507 (1989).
5. M. Neuberg, J. Adamkiewicz, J. B. Hunter, R. Müller, *Nature* **341**, 243 (1989); P. Agre, P. F. Johnson, S. L. McKnight, *Science* **246**, 922 (1989); J. W. Sellers and K. Struhl, *Nature* **341**, 74 (1989); G. Risse *et al.*, *EMBO J.* **8**, 3825 (1989).
6. Y. Nakabeppu and D. Nathans, *EMBO J.* **12**, 3833. (1989).
7. W. F. DeGrado *et al.*, *Science* **243**, 622 (1989).
8. D. C. Rees, L. DeAntonio, D. Eisenberg, *ibid.* **245**, 510 (1989); T. O. Yeates *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6438 (1987); J. U. Bowie *et al.*, *ibid.* **86**, 2152 (1989).
9. S. Marqusee and R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8898 (1987); M. Sundaralingam *et al.*, *Proteins* **2**, 64 (1987).
10. A. Zlotnick and S. L. Brenner, *J. Mol. Biol:* **209**, 447 (1989).
11. C. R. Vinson, P. B. Sigler, S. L. McKnight, *Science* **246**, 911 (1989).
12. T. Tabata *et al.*, *ibid.* **245**, 965 (1989).

13. S. R. Jordan and C. O. Pabo, *ibid.* **242**, 893 (1988); A. K. Aggarwal *et al.*, *ibid.*, p. 899; Z. Otwinowski *et al.*, *Nature* **335**, 321 (1988).
14. D. Bohmann and R. Tjian, *Cell* **59**, 709 (1989).
15. S. Y. M. Lau *et al.*, *J. Biol. Chem.* **266**, 13253 (1984); R. S. Hodges *et al.*, *ibid.* **256**, 1214 (1981).
16. Y-H. Chen, J. T. Yang, K. H. Chau, *Biochemistry* **13**, 3350 (1974).
17. N. Greenfield and G. D. Fasman, *ibid.* **8**, 4108 (1969).
18. C. N. Pace, *Methods Enzymol.* **131**, 266 (1986); J. U. Bowie and R. T. Sauer, *Biochemistry* **28**, 7139 (1989).
19. T. D. Tullius, *Annu. Rev. Biophys. Biophys. Chem.* **18**, 213 (1989).
20. W. Sanger, in *Principles of Nucleic Acid Structure*, C. R. Cantor, Ed. (Springer-Verlag, New York, 1984), p. 371.
21. W. F. DeGrado, *Adv. Prot. Chem.* **39**, 51 (1988); E. T. Kaiser and F. J. Kézdy, *Science* **223**, 249 (1984); J. W. Taylor and E. T. Kaiser, *Pharmacol. Rev.* **38**, 291 (1986).
22. A. M. Labhardt *et al.*, *Biochemistry* **22**, 321 (1983).
23. A. M. Labhardt, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7674 (1984).
24. Amino acid sequences are tabulated in M. Nishizawa *et al.*, *ibid.* **86**, 7711 (1989). 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.
25. S. Anott and D. W. L. Hukins, *Biochem. Biophys. Res. Commun.* **47**, 1504 (1972).
26. B. C. Finzel *et al.*, *J. Mol. Biol.* **186**, 627 (1985).
27. M. J. McGregor *et al.*, *ibid.* **198**, 295 (1987).
28. R. E. Bruccoleri, J. Novotny, P. Keck, C. Cohen, *Biophys. J.* **49**, 79 (1986).
29. A. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 497 (1980).
30. A. R. Oliphant, C. J. Brandl, K. Struhl, *Mol. Cell. Biol.* **9**, 2944 (1989).
31. We thank S. Jackson for peptide synthesis, J. Wendoloski and J. Krywko for expert assistance in model building, D. Janvier for amino acid analyses and oligonucleotide synthesis, C. Ampe and T. Steitz for their gift of GCN4, R. Steed for oligonucleotide purification, J. Lazaar for mass spectrometry, J. Blaney and W. Ripka for the use of a program to calculate DNA helices, and J. Novotny for sending coordinates of their model of an α-helical coiled coil. We also thank S. Brenner for helpful discussions and encouragement during the course of the work.

15 March 1990; accepted 20 June 1990.

Rapid, Sensitive Bioluminescent Reporter Technology for Naphthalene Exposure and Biodegradation

J. M. H. KING, P. M. DIGRAZIA, B. APPLEGATE, R. BURLAGE, J. SANSEVERINO, P. DUNBAR, F. LARIMER, G. S. SAYLER*

A bioluminescent reporter plasmid for naphthalene catabolism (pUTK21) was developed by transposon (Tn4431) insertion of the *lux* gene cassette from *Vibrio fischeri* into a naphthalene catabolic plasmid in *Pseudomonas fluorescens*. The insertion site of the *lux* transposon was the *nahG* gene encoding for salicylate hydroxylase. Luciferase-mediated light production from *P. fluorescens* strains harboring this plasmid was induced on exposure to naphthalene or the regulatory inducer metabolite, salicylate. In continuous culture, light induction was rapid (15 minutes) and was highly responsive to dynamic changes in naphthalene exposure. Strains harboring pUTK21 were responsive to aromatic hydrocarbon contamination in Manufactured Gas Plant soils and produced sufficient light to serve as biosensors of naphthalene exposure and reporters of naphthalene biodegradative activity. The robust and sensitive nature of the bioluminescent reporter technology suggests that new sensing methods can be developed for on-line process monitoring and control in complex environmental matrices.

BIOLUMINESCENCE IS A NATURAL phenomenon associated with several species, notably the firefly (*Photinus pyralis*) and the microorganisms *Photobacterium* and *Vibrio*. Expression and regulation of the bacterial luciferase (*lux*) genes has been studied in detail (*1*). The *lux* structural genes have been cloned and introduced into a variety of hosts (*2*). Recently, a light-emitting transcriptional fusion of the *lux* genes from *Vibrio fischeri* and the upper pathway promoter from the catabolic plasmid NAH7 was developed as a direct measure of catabolic gene induction and expression (*3*). The use of bioluminescent light as a measure of catabolic activity offers attractive applications in environmental simulations and potential field analysis for on-line analysis of microbial biodegradative activity. In comparison to conventional activity assays, bioluminescent reporter systems are noninvasive, nondestructive, rapid, and population-specific.

We describe here the construction and characterization of a bioluminescent reporter plasmid for naphthalene degradation. The application of this bioluminescent reporter system in conjunction with remote sensing technology for monitoring in situ biodegradative activity in environmental simulations was also demonstrated. A photomultiplier detection system was developed to quantify bioluminescence (*3, 4*). This type of monitoring system has distinct advantages over previous methods, such as autoradiography and scintillation counting, in that it is rapid and sensitive and it permits on-line and in situ determinations of bioluminescence.

The bioluminescent reporter plasmid (pUTK21) for naphthalene catabolism was constructed by transposon mutagenesis of a *Pseudomonas fluorescens* strain 5R with the use of the *lux* transposon Tn4431 carried on the suicide vector plasmid pUCD623 (Fig. 1A). Strain 5R is an environmental isolate from a Manufactured Gas Plant (MGP) soil. The naphthalene catabolic genes are encoded on a plasmid, pKA1, that shows homology to the archetypal naphthalene catabolic plasmid, NAH7 (*5*). In NAH7 the catabolic genes are organized into two operons, an upper pathway operon encoding for the conversion of naphthalene to salicylate and a lower pathway operon encoding for the oxidation of salicylate to acetaldehyde and pyruvate. Induction of the two operons is controlled at the transcriptional level not by naphthalene but by its metabolite, salicylate, and by the product of the regulatory gene, *nahR* (*6*).

We selected the bioluminescent construct 5RL, containing plasmid pUTK21, on the basis of its ability to produce strong inducible light when exposed to naphthalene vapor or when grown in the presence of salicylate. Biochemical and genetic mapping evidence was used to determine the probable insertion site of the *lux* transposon. Strain 5RL did not mineralize naphthalene completely and it accumulated salicylate (*7*), indicating that the insertion is in the lower-pathway operon. By restriction mapping and preliminary DNA blot hybridization analysis, we were able to localize the *lux* transposon to a 1.6-kb Pst I fragment adjacent to the *nahH* gene (*8*). On the basis of the biochemical data, and by assuming an NAH7-like organization of the lower pathway operon, we propose that the *lux* transposon inserted in the *nahG* (salicylate hydroxylase) gene (Fig. 1B).

The bioluminescent catabolic reporter plasmid pUTK21 was transferred by conjugation to another environmental *P. fluorescens* strain, strain HK9 (*9*). This strain was isolated from an MGP soil and was able to degrade salicylate but not naphthalene (Nah^-Sal^+ phenotype). The recipient strain HK44 containing pUTK21 has a Nah^+Sal^+ phenotype and exhibited the same light-producing characteristics as strain 5RL.

We evaluated the dynamic response of the

J. M. H. King, P. M. DiGrazia, B. Applegate, R. Burlage, J. Sanseverino, P. Dunbar, G. S. Sayler, Center for Environmental Biotechnology, University of Tennessee, 10515 Research Drive, Suite 200, Knoxville, TN 37932.
F. Larimer, Biology Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831.

*To whom correspondence should be addressed.