22. W. B. Curry, R. C. Thunell, S. Honjo, *Earth Planet. Sci. Lett.* **64**, 33 (1983); R. G. Fairbanks, M. Sverdlove, R. Free, P. H. Wiebe, A. W. H. Be, *Nature* **298**, 841 (1982); W. G. Deuser, E. H. Ross, C. Hemleben, M. Spindler, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **33**, 103 (1981).
23. A. Berger, *J. Atmos. Sci.* **35**, 2362 (1978).

24. We thank W. F. Ruddiman and C. Sancetta for their critical reviews. This research was supported by NSF grants OCE89-11841 and OCE85-16133. Lamont-Doherty Geological Observatory Contribution No. 4628.

27 March 1990; accepted 4 June 1990

Sequence-Specific DNA Binding by a Short Peptide Dimer

ROBERT V. TALANIAN, C. JAMES MCKNIGHT, PETER S. KIM

A recently described class of DNA binding proteins is characterized by the "bZIP" motif, which consists of a basic region that contacts DNA and an adjacent "leucine zipper" that mediates protein dimerization. A peptide model for the basic region of the yeast transcriptional activator GCN4 has been developed in which the leucine zipper has been replaced by a disulfide bond. The 34-residue peptide dimer, but not the reduced monomer, binds DNA with nanomolar affinity at 4°C. DNA binding is sequence-specific as judged by deoxyribonuclease I footprinting. Circular dichroism spectroscopy suggests that the peptide adopts a helical structure when bound to DNA. These results demonstrate directly that the GCN4 basic region is sufficient for sequence-specific DNA binding and suggest that a major function of the GCN4 leucine zipper is simply to mediate protein dimerization. Our approach provides a strategy for the design of short sequence-specific DNA binding peptides.

THE TRANSCRIPTIONAL ACTIVATOR GCN4 (*1*), which is responsible for the general control of amino acid biosynthesis in yeast (*2*), binds DNA through a structural motif common to several proteins (*3*), including the nuclear oncogene products Fos and Jun. This "bZIP" (*4*) motif consists of a region with several basic residues that probably contacts DNA directly and an adjacent region of about 30 residues containing a heptad repeat of leucines, the "leucine zipper" (*5*), that mediates dimerization. Such bZIP dimers bind DNA sites that are approximately diad-symmetric (*3*).

Structural studies of a synthetic peptide corresponding to the leucine zipper region of GCN4 indicate that the peptide dimerizes as a parallel coiled coil (*6, 7*). The leucine zipper regions are necessary for dimerization of GCN4 (*8–10*) and other bZIP proteins (*11, 12*) and for heterodimer formation by the Fos and Jun proteins (*10, 13–15*). Moreover, synthetic leucine zipper peptides are sufficient for specific homodimer (*6*) and heterodimer (*16*) formation.

The basic region of bZIP proteins is important for DNA binding. Several bZIP proteins with mutations in the basic region fail to bind DNA sequence-specifically although they can dimerize (*12, 13, 17*). Alignment of sequences from different bZIP proteins shows that conserved residues in the basic region and the leucine zipper are separated by an invariant number of residues (*4, 18*). This separation appears crucial since insertion or deletion of a few amino acid residues at the boundary between the two regions can eliminate specific DNA binding activity (*14, 19, 20*). Nevertheless, the two regions appear capable of functioning autonomously, since chimeric bZIP domains (combining the basic region of one protein with the leucine zipper of another) often retain specific DNA binding activity (*10, 20, 21*).

We asked whether the basic region alone, dimerized with a disulfide in place of the leucine zipper, retains sequence-specific DNA binding activity. A peptide (GCN4-brl), corresponding to residues 222 to 252 of GCN4 (*22*), was synthesized (*23*) with a Gly-Gly-Cys linker (*6*) added at the carboxyl terminus (Fig. 1). The glycines were included to provide a flexible linker in the disulfide-bonded dimer, referred to as GCN4-brl^{ss}. The peptide was made as the carboxyl-terminal amide to avoid introduction of additional charge. A second peptide (GCN4-bZIP1), corresponding to the entire bZIP region of GCN4 (residues 222 to 281), was also synthesized (Fig. 1). This 60-residue peptide is capable of dimerization and sequence-specific DNA binding (*8*).

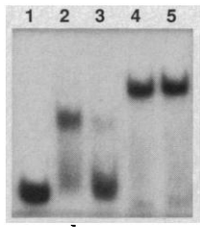

Fig. 2. Gel mobility shift assays (*24*) indicate that DNA binding by GCN4-brl^{ss}, but not GCN4-bZIP1, is sensitive to DTT. Lane 1, no peptide; lane 2, GCN4-brl^{ss}; lane 3, GCN4-brl^{ss} with 10 mM DTT; lane 4, GCN4-bZIP1; and lane 5, GCN4-bZIP1 with 10 mM DTT.

Gel mobility shift assays (*24*) indicate (Fig. 2) that both GCN4-brl^{ss} and GCN4-bZIP1 bind a 20-bp oligonucleotide, GRE20 (*24*), which contains the GCN4 recognition element (GRE) 5′-ATGACTCAT-3′ (*25*). As measured by titration of the gel shift, GCN4-brl^{ss} binds GRE20 with a dissociation constant of ~10 nM at 4°C. Reduction of the disulfide bond in GCN4-brl^{ss} by addition of 10 mM dithiothreitol (DTT) decreases substantially the amount of mobility-shifted DNA, whereas DNA binding by GCN4-bZIP1 is unaffected by this treatment (Fig. 2).

The DNA binding specificities of GCN4-brl^{ss} and GCN4-bZIP1 were tested by using deoxyribonuclease (DNase) I footprinting (*26*). At 4°C both peptides show sequence-specific protection of the GRE site from DNase I digestion (Fig. 3). However, when DNase I digestion was carried out at 24°C, GCN4-brl^{ss} failed to bind specifically, although GCN4-bZIP1 gave an identical footprint to that obtained at 4°C.

The DNA binding specificity of GCN4-brl^{ss} suggests that the peptide is a valid model for the DNA binding activity of GCN4. The binding activity of the peptide dimer demonstrates directly that the basic region of GCN4 (and presumably other bZIP proteins) contains sufficient information for sequence-specific DNA binding. The successful substitution of the leucine

Fig. 1. Sequences of the peptides studied (*23*). GCN4-bZIP1 consists of the 60 carboxyl-terminal residues of GCN4 (*22*). The leucines in the leucine repeat are underlined. GCN4-brl consists of the basic region residues (boxed) plus the carboxyl-terminal linker Gly-Gly-Cys. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; G, Gly; H, His; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; and Y, Tyr.

Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142, and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

zipper with a flexible disulfide linker, and the dependence of DNA binding on the presence of the disulfide bond, suggest that the primary function of the leucine zipper is dimerization. However, DNA binding by GCN4-brl^{ss}, but not GCN4-bZIP1, is temperature dependent between 4° and 24°C. These observations suggest an additional role for the leucine zipper [for example, orientation of the DNA binding regions; see (*14, 19, 20*)] that is not modeled by the flexible disulfide linker.

Structural studies can be simplified by using peptide models for protein motifs. Accordingly, we have used circular dichroism (CD) spectroscopy to examine the secondary structure of GCN4-brl^{ss} in the presence and absence of GRE20. The CD spectrum of the peptide (Fig. 4A) suggests that it shows partial α-helix formation in the absence of DNA (*27*). The intensity of the CD signal of GCN4-brl^{ss} at 222 nm (a helical band) increases substantially upon addition of an equimolar amount of GRE20 (Fig. 4B). The small change in the region of the spectrum dominated by signals from the oligonucleotide (245 to 310 nm) suggests that the much larger changes observed below 245 nm result primarily from changes in peptide rather than oligonucleotide structure. The difference spectrum (Fig. 4C) indicates that the peptide is highly α helical when bound to DNA (*28*). These results are consistent with both the "scissors grip" (*4*) and "induced helical fork" (*29*) models, which postulate that the basic regions of bZIP proteins bind DNA in an α-helical conformation.

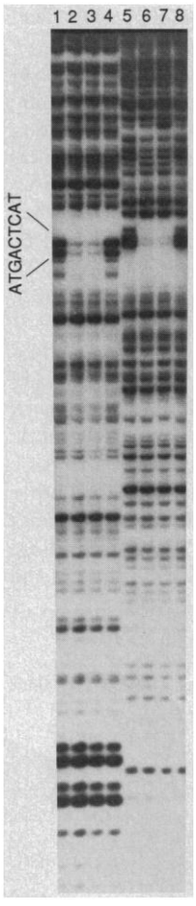

Fig. 3. The DNase I footprint of the GCN4 binding site by GCN4-brl^{ss} and GCN4-bZIP1 are identical (*26*). Lanes 1 to 4, DNA labeled with ^{32}P-phosphate at the 5′ end of the ⊕ strand; lanes 5 to 8, label on the ⊖ strand; lanes 1, 4, 5, and 8, DNase I control (no peptides present); lanes 2 and 6, GCN4-brl^{ss}; and lanes 3 and 7, GCN4-bZIP1.

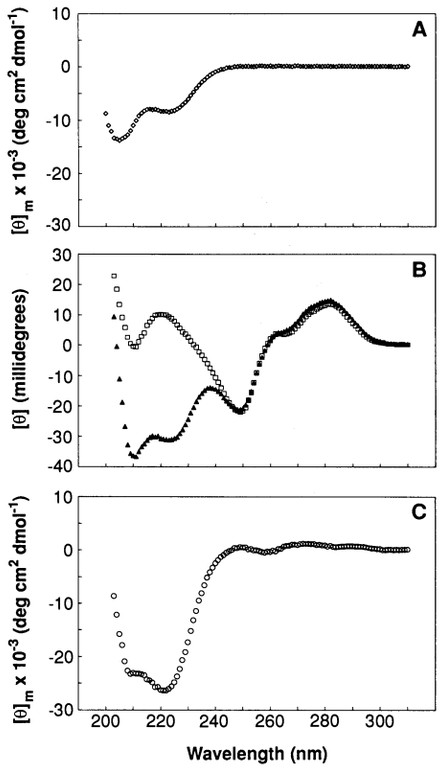

Fig. 4. CD difference spectroscopy indicates that GCN4-brl^{ss} is helical when bound to DNA (*32*). (**A**) GCN4-brl^{ss} alone. (**B**) GRE20 alone (□) and GCN4-brl^{SS} with GRE20 (▲). (**C**) Spectrum of GCN4-brl^{ss} bound to GRE20 calculated as the difference between the two spectra in (B).

Although GCN4-brl^{ss} is a remarkably short DNA binding peptide, it seems likely that even shorter peptides with sequence-specific DNA binding activity can be made. For example, several of the amino-terminal residues in the basic region used here have been found recently to be dispensable for DNA binding (*19, 30*). In addition, the use of a Gly-Gly-Cys (*6*) or other linker [see, for example, (*31*)] could lead to peptide models for other DNA binding motifs. Peptide models like GCN4-brl^{ss} hold promise for structural studies of sequence-specific protein-DNA interactions and for the design of short, sequence-specific DNA binding peptides.

REFERENCES AND NOTES

1. M. D. Penn, B. Galgoci, H. Greer, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2704 (1983); A. G. Hinnebusch and G. R. Fink, *ibid.*, p. 5374.
2. E. W. Jones and G. R. Fink, in *The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression*, J. N. Strathern, E. W. Jones, J. R. Broach, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982), pp. 181–299.
3. Reviewed by P. F. Johnson and S. L. McKnight, *Annu. Rev. Biochem.* **58**, 799 (1989); K. Struhl, *Trends Biochem. Sci.* **14**, 137 (1989).
4. C. R. Vinson, P. B. Sigler, S. L. McKnight, *Science* **246**, 911 (1989).
5. W. H. Landschultz, P. F. Johnson, S. L. McKnight, *ibid.* **240**, 1759 (1988).
6. E. K. O'Shea, R. Rutkowski, P. S. Kim, *ibid.* **243**, 538 (1989).
7. T. G. Oas, L. P. McIntosh, E. K. O'Shea, F. W. Dahlquist, P. S. Kim, *Biochemistry* **29**, 2891 (1990); R. Rasmussen, D. Benvegnu, E. K. O'Shea, P. S. Kim, T. Alber, unpublished results.
8. I. A. Hope and K. Struhl, *Cell* **46**, 885 (1986).
9. ———, *EMBO J.* **6**, 2781 (1987).
10. T. Kouzarides and E. Ziff, *Nature* **340**, 568 (1989); J. W. Sellers and K. Struhl, *ibid.* **341**, 74 (1989).
11. C. V. Dang, M. McGuire, M. Buckmire, W. M. F. Lee, *ibid.* **337**, 664 (1989).
12. W. H. Landschultz, P. F. Johnson, S. L. McKnight, *Science* **243**, 1681 (1989); V. J. Dwarki, M. Montminy, I. M. Verma, *EMBO J.* **9**, 225 (1990).
13. R. Turner and R. Tjian, *Science* **243**, 1689 (1989); R. Gentz, F. J. Rauscher III, C. Abate, T. Curran, *ibid.*, p. 1695; M. Neuberg, M. Schuermann, J. B. Hunter, R. Müller, *Nature* **338**, 589 (1989).
14. T. Smeal, P. Angel, J. Meek, M. Karin, *Genes Dev.* **3**, 2019 (1989).
15. L. J. Ransone, J. Visvander, P. Sassone-Corsi, I. M. Verma, *ibid.*, p. 770; M. Schuermann *et al.*, *Cell* **56**, 507 (1989).
16. E. K. O'Shea, R. Rutkowski, W. F. Stafford III, P. S. Kim, *Science* **245**, 646 (1989).
17. D. Bohmann and R. Tjian, *Cell* **59**, 709 (1989).
18. P. K. Vogt, T. J. Bos, R. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3316 (1987).
19. K. Struhl, personal communication.
20. P. Agre *et al.*, *Science* **246**, 922 (1989).
21. K. Struhl, *Cell* **50**, 841 (1987); Y. Nakabeppu and D. Nathans, *EMBO J.* **8**, 3833 (1989); M. Neuberg, J. Adamkiewicz, J. B. Hunter, R. Müller, *Nature* **341**, 243 (1989).
22. G. Thireos, M. D. Penn, H. Greer, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5096 (1984); A. G. Hinnebusch, *ibid.*, p. 6442.
23. Peptides were synthesized on an Applied Biosystems Model 430A peptide synthesizer with standard reaction cycles modified to include acetic anhydride capping. Peptides were cleaved from the resins by low-high HF cleavage (Immunodynamics, Inc., San Diego, CA) and desalted by Sephadex G-10 chromatography in 5% acetic acid. Purifications were by high-performance liquid chromatography with a Vydac reverse-phase C_{18} column and a linear gradient of CH_3 CN-H_2O with 0.1% trifluoroacetic acid. Fast atom bombardment mass spectrometry: GCN4-brl: calculated, 3796.5; found, 3795.8; GCN4-bZIP1: calculated, 7015.4; found, 7015.5.
24. M. M. Garner and A. Revzin, *Nucleic Acids Res.* **9**, 3047 (1981); M. Fried and D. M. Crothers, *ibid.*, p. 6505. Binding solutions contained in 15 μl: 20 mM tris (pH 7.4), 4 mM KCl, 2 mM $MgCl_2$, 2 mM EDTA, 0.1% NP-40 detergent, 5000 cpm (~2 fmol) 5′-^{32}P–labeled GRE20 (5′-CGGATGACTCAT-TTTTTTTC-3′, double-stranded) and 250 nM peptides when present. Mixtures were incubated at 4°C for 30 min. Free and peptide-bound labeled GRE20 were resolved by nondenaturing 8% polyacrylamide gel electrophoresis in TE buffer at 4°C. Peptide concentrations were determined by tyrosine absorbance [H. Edelhoch, *Biochemistry* **6**, 1948 (1967)] or by quantitative amino acid analysis, and DNA concentration was determined with an extinction coefficient at 260 nm (ϵ_{260}) of 473,000 M^{-1} cm^{-1} [R. B. Wallace and C. G. Miyada, *Methods Enzymol.* **152**, 423 (1987)].
25. D. E. Hill, I. A. Hope, J. P. Macke, K. Struhl, *Science* **234**, 451 (1986).
26. D. J. Galas and A. Schmitz, *Nucleic Acids Res.* **5**, 3157 (1978). Solutions (4°C) contained in 200 μl: 20 mM tris (pH 7.4), 4 mM KCl, 2 mM $MgCl_2$, 0.1% NP-40, 200 μM (bp) sonicated salmon sperm DNA, 15,000 cpm (~5 fmol) of DNA probe and 1.0 μM GCN4-brl^{ss} and 1.1 μM GCN4-bZIP1 when present. DNA probes were prepared by poly-

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.