404 (1985).
3. ———, J. Voyvodic, *J. Neurosci.* **6**, 1051 (1986).
4. L. Magrassi, D. Purves, J. W. Lichtman, *ibid.* 7, 1207 (1987).
5. J. W. Lichtman, L. Magrassi, D. Purves, *ibid.*, p. 1215.
6. D. Purves and J. W. Lichtman, *ibid.*, p. 1492.
7. W. D. Snider, *ibid.*, p. 1760.
8. J. W. Lichtman, *J. Physiol. (London)* **273**, 155 (1977).
9. J. Voyvodic, *Soc. Neurosci. Abstr.* **12**, 390 (1986).
10. R. Hume and D. Purves, *J. Physiol. (London)* **338**, 259 (1983).
11. C. J. Forehand and D. Purves, *J. Neurosci.* **4**, 1 (1984).
12. Correlation of pairs of composite maps on a pixel by pixel basis was not routinely done because such a detailed analysis emphasizes apparent differences in the two maps that are, in fact, below the resolving power of our imaging technique.
13. C. W. Cotman, M. Nieto-Sampedro, E. W. Harris, *Physiol. Rev.* **61**, 684 (1981); M. C. Brown, R. L. Holland, W. G. Hopkins, *Annu. Rev. Neurosci.* **4**, 17 (1981); L. Mendell, *Physiol. Rev.* **64**, 260 (1984); D. Purves and J. W. Lichtman, *Principles of Neural Development* (Sinauer, Sunderland, MA, 1985).
14. A. Wernig and A. A. Herrera, *Prog. Neurobiol.* **27**, 251 (1986); A. A. Herrera and L. R. Banner, *Soc. Neurosci. Abstr.* **13**, 1665 (1987).
15. O. Ostle and R. W. Mensing, *Statistics in Research* (Iowa State Univ. Press, Ames, ed. 3, 1975), p. 237.
16. We thank W. J. Sunderland for his expert help with various aspects of the image processing and graphic analysis procedures and C. J. Forehand, L. Harris, S. L. Pomeroy, and J. R. Sanes for helpful comments on the manuscript.

6 May 1987; accepted 6 October 1987

A New Prosomatostatin-Derived Peptide Reveals a Pattern for Prohormone Cleavage at Monobasic Sites

ROBERT BENOIT,* NICHOLAS LING, FREDERICK ESCH

Cleavage of the peptide bonds of preprosomatostatin at basic residues near the carboxyl terminus yields somatostatin-14, somatostatin-28, and somatostatin-28(1-12). However, little is known about the molecular forms derived from the amino terminal portion of the precursor, even though this part of the prohormone is highly conserved through evolution. By using an antibody against the amino terminus of prosomatostatin, a decapeptide with the structure Ala-Pro-Ser-Asp-Pro-Arg-Leu-Arg-Gln-Phe, corresponding to preprosomatostatin(25-34), was isolated from the endocrine portion of the rat stomach, the gastric antrum. The antral decapeptide may represent a bioactive product generated from prosomatostatin after a monobasic cleavage similar to that involved in the formation of somatostatin-28. In fact, a monobasic cleavage requires two basic residues and a domain containing nonpolar amino acids such as alanine or leucine, or both.

PEPTIDES DESTINED FOR SECRETION from cells are derived from precursor proteins by post-translational modifications. Most important among these modifications are the enzymatic cleavages that usually occur at some dibasic residues and occasionally at monobasic sites. The primary structure of neurohormone precursors can be elucidated rapidly by recombinant DNA technology. It then becomes important to predict how these precursors are processed in order to identify the new peptide products derived from the prohormones.

The somatostatin precursor is a useful model for studying neuropeptide processing in mammals, largely because of its simplicity. A single gene codes for the 92–amino acid prohormone (*1*). The precursor is not glycosylated, and contains only one pair of basic amino acids (preceding somatostatin-14) and no repetitive sequences. The region of the prohormone endowed with the pan-inhibitory activity corresponds to the cyclic structure formed by its last 12 amino acids (*2*). Since the identification of the somatostatin precursor, we have been searching for additional bioactive domains contained in the 80–amino acid sequence preceding the carboxyl terminal dodecapeptide. Using antibody probes directed toward the carboxyl terminus of the precursor, we have identified two peptides derived from prosomatostatin (*3, 4*) that do not contain the original tetradecapeptide somatostatin: somatostatin-28(1-12) and preprosomatostatin(25-100). Even though these two peptides have been established as secretory products and are widely distributed in brain and digestive system, no biological role has been ascribed to them (*5*).

The presence of somatostatin-28 and somatostatin-28(1-12) in tissue extracts indicates that classic cleavage at a pair of basic amino acids is not the sole type of peptide hydrolysis during processing of prosomatostatin (*6, 7*). Cleavage at the monobasic site Arg^{88}-Ser^{89} also occurs (Fig. 1). Recently, efforts have been directed toward the identification of putative peptides derived from the amino terminus of prosomatostatin (*4, 5, 8, 9*). Using an antibody directed toward the preprosomatostatin(25-33) segment, we have screened tissues of adult rats for the presence of preprosomatostatin(25-33)–like material (*10*) with a radioimmunoassay (RIA). The hypothalamus and the stomach contained the highest concentration of immunoreactive material (9.1 ± 1.4 and 6.2 ± 0.5 ng per 100 mg of wet weight, respectively; mean ± SEM; $n = 10$ rats). When this material was subjected to gel permeation chromatography, the stomach was found to contain an immunoreactive 1.2- to 1.4-kD species that was virtually absent in brain, pancreas, and duodenum. Thus differential processing occurred at the amino terminus of prosomatostatin. Here we describe the isolation and characterization of the 1.4-kD peptide. In addition, we propose a new processing scheme to account for the derivation of the peptide from prosomatostatin.

Stomachs from three Sprague-Dawley rats were collected and the corpus, antrum, and cutaneous region dissected (*11*) and extracted in 2*M* acetic acid containing peptidase inhibitors. RIA measurements were performed on the extract or on the fractions eluted from Sephadex G-75 gel permeation chromatography columns (Fig. 2).

Immunoreactivity in the cutaneous zone was low (total content per stomach, 2 ng). The corpus of the stomach contained 18.5 ng of preprosomatostatin(25-33)–like immunoreactive material of which 14.3 ng eluted from the gel permeation column as large molecular forms and 4.2 ng as a molec-

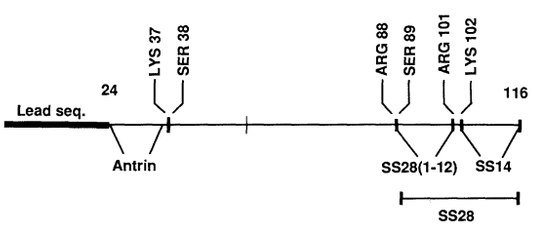

Fig. 1. Schematic representation of 116–amino acid mammalian preprosomatostatin, which includes the tetradecapeptide somatostatin-14 (SS14), somatostatin-28 (SS28), somatostatin-28(1-12) [SS28(1-12)], the amino terminal peptide antrin, and the 24–amino acid leader sequence. The three main cleavage points of the prohormone at positions 101-102, 88-89, and, conceivably, 37-38 are shown.

R. Benoit, Montreal General Hospital Research Institute, Department of Medicine, McGill University, Montreal, Quebec H3G 1A4, Canada.
N. Ling and F. Esch, Laboratories for Neuroendocrinology, Salk Institute, La Jolla, CA 92138.

*To whom correspondence should be addressed.

ular form of about 1.4 kD (Fig. 2). In contrast, the rat antrum contained only 4.2 ng of immunoreactive material eluting as large molecular forms and 18.5 ng in the 1.4-kD peptide zone. The product ratio of the small versus large forms was thus 15 times higher in antrum than in corpus. Modifications of the acid concentration or peptidase inhibitors during extraction or omission of the heating step yielded essentially the same chromatography profile. To further ascertain that the 1.4-kD molecular form was not generated during extraction, a total of 415 ng of immunoreactive material was obtained by pooling fractions from gel filtration of a gastric corpus extract that corresponded to peptides between 6 and 10 kD. This pool was added to antral tissue before extraction and gel filtration. No additional 1.4-kD peptide was generated by this procedure. Thus this new molecular form represents an authentic endogenous product. Because its concentration was 88 ng per gram of wet weight in the gastric antrum, isolation of the 1.4-kD peptide was carried out from antral tissue.

The antrums of 378 Sprague-Dawley rats were frozen in liquid nitrogen immediately after dissection and lyophilized. The tissue was extracted with acid and peptidase inhibitors (Fig. 2). The extract (3.4 g dry weight) was first purified by gel permeation on Sephadex G-75 and eluted with 5.2*M* acetic acid. The immunoreactive material detected in the 1.4-kD zone from several columns was pooled to yield a total of 52 mg of dry material containing 6.7 μg of preprosomatostatin(25-33)–like peptide. The second purification step was ion-exchange chromatography on carboxymethyl cellulose (CMC). Effective separation was performed by using a constant *p*H of 4.6 and a shallow gradient of ammonium acetate, so that 6.5 μg of the 1.4-kD peptide was recovered (*12*). Further purification was carried out by analytical reversed-phase high-performance liquid chromatography (RPLC), with octadecylsilyl-silica (ODS) as the solid phase (*13*). The mobile phase consisted of heptafluorobutyric acid–acetonitrile or, as was the case for the final RPLC purification (Fig. 3), trifluoroacetic acid (TFA)–acetonitrile (*14*). With this approach, 352 pmol of homogeneous 1.4-kD immunoreactive peptide were isolated. We sequenced 220 pmol of the peptide with an Applied Biosystem model 470A gas-phase sequenator (*15*). The initial yield was 97.4%, and the average repetitive yield was 92.5%. The primary structure determined was Ala-Pro-Ser-Asp-Pro-Arg-Leu-Arg-Gln-Phe. Duplicate amino acid analyses (*16*) performed on 8 pmol of material revealed the following composition: Asx, 1.13 ± 0.04; Ser, 0.98 ± 0.00; Glx, 1.05 ± 0.03; Ala, 0.93 ± 0.01; Leu, 0.93 ± 0.03; Phe, 0.96 ± 0.03; Arg, 1.85 ± 0.01; Pro, 2.16 ± 0.04; Cys, Lys, Trp, His, Tyr, Ile, Met, Val, Gly, and Thr, 0 (mean ± SD, $n = 2$). This composition confirmed that the antral peptide is a decapeptide of 1186 daltons.

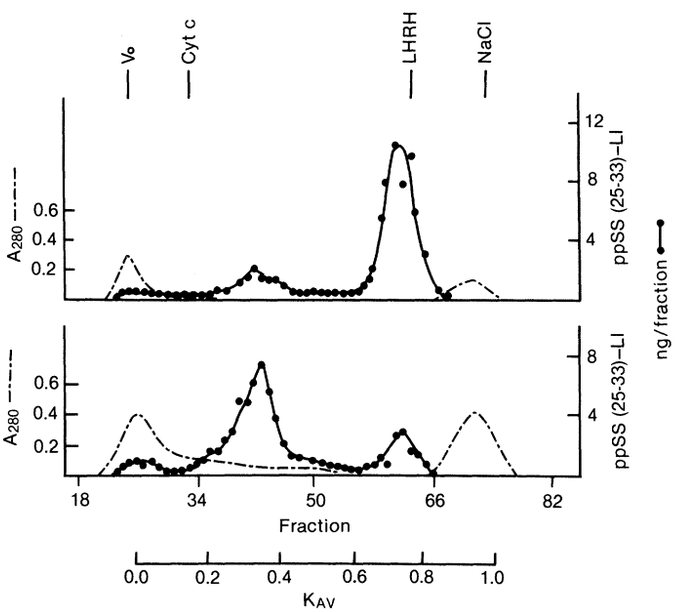

Fig. 2. Sephadex G-75 gel permeation chromatography performed on extracts of two gastric regions. Three male Sprague-Dawley rats (310 g) that had been allowed free access to food were killed by decapitation; the antrum and the corpus of the stomach were dissected, washed in 150 m*M* NaCl, and heated at 95°C for 12 minutes in extraction buffer [2*M* acetic acid containing pepstatin (20 mg/liter), phenylmethylsulfonyl fluoride (20 mg/liter), and bacitracin (50 mg/liter)]. The three antrums and the three corpora were homogenized separately in 8 ml of extraction buffer per gram of tissue with a polytron (Brinkmann). The homogenate was centrifuged at 40,000*g* at 2°C for 20 minutes and the supernatant lyophilized. The corpus extract (130 mg dry weight) and the antral extract (42 mg dry weight) were dissolved in 5 ml of 5.2*M* acetic acid in water and loaded separately onto a column (2.5 by 100 cm) packed with Sephadex G-75 (Pharmacia). Elution was done at 4°C with 5.2*M* acetic acid at a flow rate of 23 ml/hour. Fraction size was 7 ml and preprosomatostatin(25-33)–like immunoreactivity [ppSS(25-33)-LI] was measured in each fraction. The column had been calibrated with bovine serum albumin (Vo), 67 kD; cytochrome c(Cyt c), 12.4 kD; and luteinizing hormone–releasing hormone (LHRH), 1182 daltons. The immunoreactive profile obtained after fractionation of the antral extract is shown in the upper panel. The lower panel shows the immunoreactive profile obtained after fractionation of the corpus extract. A_{280}, absorbance at 280 nm.

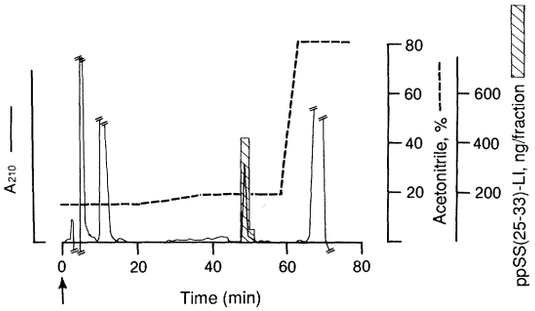

Fig. 3. Final purification by RPLC of the preprosomatostatin(25-33)–like immunoreactivity [ppSS(25-33)-LI] from rat antrum. Loading was done in 1 ml of 0.1% TFA in water; 352 pmol of peptide were eluted isocratically after 12 minutes at 20% acetonitrile in 0.1% TFA. Fraction size was 1.2 ml, flow rate was 0.6 ml/min, and the spectrophotometer sensitivity was 0.04.

The antral peptide, or antrin, is preprosomatostatin(25-34). Previous work suggested that a larger peptide of 32 amino acids would be derived from the amino terminal portion of rat preprosomatostatin (*4*). This peptide, obtained after cleavage at the sequence Leu^{56}-Leu^{57} of the precursor has been identified by Schmidt *et al.* (*9*) in extracts of porcine intestine. However, cleavage of this Leu-Leu bond leads to only minute quantities of this product (*5*). Thus, antrin may represent the most interesting peptide derived from the amino terminus of prosomatostatin in the murine digestive system. A biological role for rat antrin is suggested by its preferential distribution in one digestive organ, particularly in the endocrine portion of this organ, the gastric antrum. In addition, the preprosomatostatin(25-34) sequence that corresponds to antrin is highly conserved from catfish to man (*17*).

The generation of the antral decapeptide provides some insight on the processing of prosomatostatin and other prohormones. The amino acid sequence of antrin is followed by the sequence Leu–Gln–basic residue–Ser, which is the same sequence as that involved in the formation of somatostatin-28 and somatostatin-28(1-12). We propose that the peptide bond Lys^{37}-Ser^{38} is cleaved similarly to the other monobasic site Arg^{88}-Ser^{89} (Fig. 4). Formation of the mature ten–

amino acid product antrin could then involve the action of a carboxypeptidase. In spite of this structural similarity, the conditions required for cleavage at the two monobasic sites are different, because large amounts of somatostatin-28 are present in duodenum (*18*), whereas only very low quantities of the decapeptide antrin occur in that region.

Until recently, the only prohormone known to be processed at a monobasic site was pro–arginine vasopressin (*19*). Now it has been established that the rat arginine vasopressin precursor, as well as the pro–egg laying hormone of *Aplysia* (*20*), the human pro–growth hormone releasing factor-40 (*21*), and procholecystokinin-58 (*22*), are all cleaved at an Arg-Ala bond (Fig. 4). Two characteristics of this cleavage are that (i) the basic amino acid involved is located usually in a region rich in alanine or leucine, or both, and (ii) an arginine usually precedes the basic amino acid involved in this cleavage (Fig. 4, arrows). Surprisingly, this arginine usually occurs three or five residues before the basic amino acid involved in the cleavage, suggesting that some unknown but precise configuration is a prerequisite for cleavage. Although prosomatostatin contains ten single basic amino acids, only those sharing these two important characteristics are cleaved, that is, residues 37 and 88.

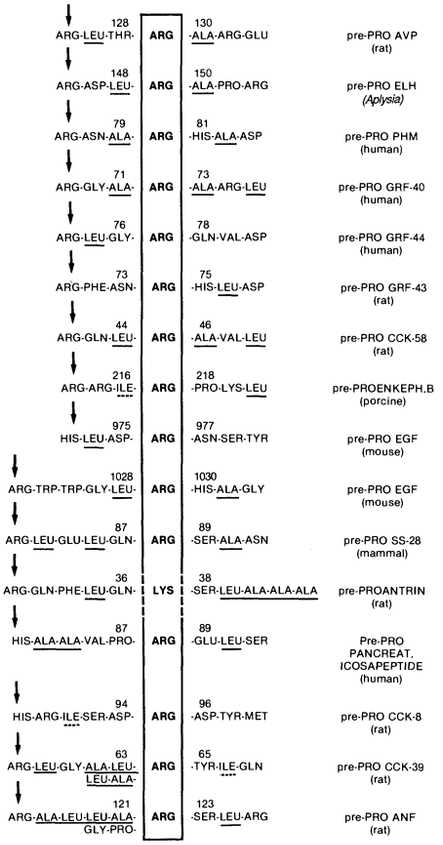

Fig. 4. Partial sequences of preprohormones known to be cleaved at monobasic sites: rat pre–pro–arginine vasopressin (AVP) (*19*), *Aplysia* pre–pro–egg-laying hormone (ELH) (*20*), human pre–pro–peptide histidyl methionine (PHM) (*29*), human pre–pro–growth hormone–releasing factors (GRF)-40 and -44 (*21*), rat pre–pro–growth hormone–releasing factor-43 (*30*), rat preprocholecystokinin (CCK)-58 (*22*), porcine preproenkephalin B or dynorphin-8 (*27*), mouse pre–pro–epidermal growth factor (EGF) (*24*), mammalian preprosomatostatin (SS)-28 (*1*), rat preproantrin, human pre–pro–pancreatic icosapeptide (*31*), rat preprocholecystokinins-8 and -39 (*22*), and pre–pro–atrial natriuretic peptide (ANF) (*28*). The basic amino acids being cleaved at their carboxyl end are boxed. The arrow indicates the preceding arginine (histidine) three, five, or seven amino acids toward the amino terminus. The hydrophobic amino acids alanine and leucine (isoleucine) are underlined.

This processing model may explain the existence in human tissue of both growth hormone–releasing factor-40 (GRF-40) and GRF-44 amide (*23*). Both gastric inhibitory peptide (GIP) and epidermal growth factor (EGF) are generated from their respective precursor after cleavage at monobasic sites immediately preceding and following the mature post-translational product (*24*). This situation is unexpected in the case of proEGF, because this precursor also contains 12 basic pairs susceptible to cleavage.

Schwartz recently concluded that there is no consensus sequence of amino acids surrounding the single basic residue—often an arginine—at which a cleavage occurs (*25*). However, he proposed that a proline-directed arginyl cleavage is involved in about one-third of the cleavages. We think that if a proline-directed arginyl cleavage exists at all, it is an exception rather than the rule. Instead we propose that three principles govern monobasic cleavages occurring in hormone and neuropeptide precursors: (i) The monobasic cleavage occurs virtually always at the carboxyl terminal end of an arginine; preproantrin and preproFMRF-amide (*26*) are rare exceptions where cleavage would occur at a lysine. (ii) A leucine (isoleucine) or several alanines are virtually always present in the two amino acids immediately preceding and the two amino acids following the monobasic cleavage. If a glycine is present, often along with a valine, as is the case for GRF-44, FMRF-amide, and adrenorphin (*27*), then amidation should occur. (iii) Three or five or seven amino acids before the arginine, a second arginine is present; rarely, this second arginine will be replaced by a histidine. Thus a so-called monobasic cleavage requires two basic residues that may be as much as seven residues apart separated by a nonpolar sequence such as Ala-Leu-Leu-Ala, as is the case in preprocholecystokinin-39 and pre–pro–atrial natriuretic peptide (*22, 28*). Such nonpolar sequences appear to play an important role in the processing at monobasic sites. The mammalian preprosomatostatin(39-42) sequence Leu-Ala-Ala-Ala (Fig. 4) probably represents such a nonpolar marker. We think that these three principles, as well as the existence of nonpolar markers, may help predict the sites of monobasic cleavages in precursors.

REFERENCES AND NOTES

1. L.-P. Shen, R. L. Pictet, W. J. Rutter, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4575 (1982); R. H. Goodman, D. C. Aron, B. A. Roos, *J. Biol. Chem.* **258**, 5570 (1983); C. L. Funckes *et al.*, *ibid*, p. 8781.
2. P. Brazeau *et al.*, *Science* **179**, 77 (1973); W. Vale *et al.*, *Recent Prog. Horm. Res.* **31**, 365 (1975).
3. R. Benoit *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 917 (1982).
4. R. Benoit, P. Böhlen, F. Esch, N. Ling, *Brain Res.* **311**, 23 (1984).
5. R. Benoit, in *Proceedings of the International Conference on Somatostatin*, S. Reichlin, Ed. (Plenum, New York, 1987), pp. 33–50; R. Benoit *et al.*, *Adv. Exp. Med. Biol.* **188**, 89 (1985).
6. L. Pradayrol, H. Jörnvall, V. Mutt, A. Ribet, *FEBS Lett.* **109**, 55 (1980); A. V. Schally *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 77, 4489 (1980); F. Esch *et al.*, *ibid.*, p. 6827; J. Spiess, J. Villarreal, W. Vale, *Biochemistry* **20**, 1982 (1981); P. Böhlen, P. Brazeau, F. Esch, N. Ling, R. Guillemin, *Regul. Pept.* **2**, 359 (1981); Y. Arakawa and S. Tachibana, *Life Sci.* **35**, 2529 (1984).
7. C. Patzelt, H. S. Tager, R. J. Carroll, D. F. Steiner, *Proc. Natl. Acad. Sci. U.S.A.* 77, 2410 (1980); P. Gluschankof *et al.*, *ibid.* **81**, 6662 (1984).
8. D. C. Aron and B. A. Roos, *Endocrinology* **118**, 218 (1986).
9. W. E. Schmidt *et al.*, *FEBS Lett.* **192**, 141 (1985).
10. A peptide corresponding to preprosomatostatin(25-33) with an additional tyrosine at the carboxyl terminus was synthesized by solid phase [N. Ling *et al.*, *Biochem. Biophys. Res. Commun.* **95**, 945 (1980)]. The synthetic peptide was coupled to bovine serum albumin with bis-diazotized benzidine, and the conjugate was injected into New Zealand White rabbits. An immune plasma was generated (SMo5) and used at a dilution of 1:1100 for RIA. Preprosomatostatin(25-33)-Tyr was labeled with chloramine-T, the ^{125}I-labeled peptide was purified on CMC and used in RIAs [R. Benoit, N. Ling, P. Brazeau, S. Lavielle, R. Guillemin, in *Neuromethods: Peptides*, A. A. Boulton, G. B. Baker, Q. J. Pittman, Eds. (Humana Press, Clifton, NJ, 1987), vol. 6, pp. 43–72]. Preprosomatostatin(25-33)-Tyr was the unlabeled ligand. Incubation was carried out in 400 μl of RIA buffer at 4°C for 6 hours and the antigen-antibody complex precipitated with 2 ml of 99% ethanol. Sensitivity was 15 fmol per RIA tube.
11. R. Hebel and M. W. Stromberg, in *Anatomy of the Laboratory Rat* (Williams and Wilkins, Baltimore, 1976), p. 47.
12. The preprosomatostatin(25-33)–like material of 1.4 kD obtained from six runs on Sephadex G-75 gel permeation chromatography columns was further purified by chromatography on CMC. The dry material obtained after the gel permeation and lyophilization was dissolved in 15 ml of 0.01*M* ammonium acetate, *p*H 4.6, and loaded onto a CM-52 column (Whatman, 1 by 16 cm). Elution was at 4°C with a linear gradient of 0.01*M* ammonium acetate to 0.06*M* at a constant *p*H of 4.6, followed by a 0.9*M* ammonium acetate wash. Flow rate was 12 ml/hour, and the fraction size was 4 ml.
13. CMC fractions containing the immunoreactive material were pooled, dried on a Speed Vac concentrator, and subjected to RPLC on a Beckman model 420 gradient liquid chromatograph in conjunction with an Altex C18 column (4.6 by 150 mm, 5-μm particle size) and a Shimadzu SPD-6A variable wavelength ultraviolet detector. The first mobile phase used was 0.12% heptafluorobutyric acid with an acetonitrile gradient of 20 to 29.6% over 80

minutes. The immunoreactive material was eluted in one peak at 26.8% acetonitrile. This peak was then chromatographed on a final RPLC run with 0.1% TFA-acetonitrile as the mobile phase.
14. H. P. J. Bennett, C. A. Browne, S. Solomon, *J. Liq. Chromatogr.* **3**, 1353 (1980).
15. F. Esch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6507 (1985).
16. P. Böhlen and R. Schroeder, *Anal. Biochem.* **126**, 144 (1982).
17. P. Argos, W. L. Taylor, C. D. Minth, J. E. Dixon, *J. Biol. Chem.* **258**, 8788 (1983).
18. Y. C. Patel, H. H. Zingg, C. B. Srikant, *Adv. Exp. Med. Biol.* **188**, 71 (1985); D. F. Trent and G. C. Weir, *Endocrinology* **108**, 2033 (1981).
19. D. Richter, in *The Posterior Pituitary: Hormone Secretion in Health and Disease*. P. H. Baylis and P. L. Padfield, Eds. (Dekker, New York, 1985), pp. 37–51.
20. R. Newcomb and R. H. Scheller, *J. Neurosci.* 7, 854 (1987).
21. U. Gubler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4311 (1983).
22. R. J. Deschenes *et al.*, *ibid.* **81**, 726 (1984).
23. R. Guillemin *et al.*, *Science* **218**, 585 (1982); J. Spiess, J. Rivier, W. Vale, *Nature (London)* **303**, 532 (1983).
24. J. Takeda *et al.*, *Endocr. Soc. Abstr.* 1011 (1987); A. Gray, T. J. Dull, A. Ullrich, *Nature (London)* **303**, 724 (1983).
25. T. W. Schwartz, *FEBS Lett.* **200**, 1 (1986).
26. M. Schaefer *et al.*, *Cell* **41**, 457 (1985).
27. H. Matsuo, A. Miyata, K. Mizuno, *Nature (London)* **305**, 721 (1983).
28. G. Thibault *et al.*, *FEBS Lett.* **167**, 352 (1984); M. Yamanaka *et al.*, *Nature (London)* **309**, 719 (1984); M. Maki *et al.*, *ibid*, p. 722; T. G. Flynn, M. L. deBold, A. J. deBold, *Biochem. Biophys. Res. Commun.* **117**, 859 (1983).
29. N. Itoh, K. Obata, N. Yanaihara, H. Okamoto, *Nature (London)* **304**, 547 (1983).
30. K. E. Mayo, G. M. Cerelli, M. G. Rosenfeld, R. M. Evans, *ibid.* **314**, 464 (1985).
31. E. Boel, T. W. Schwartz, K. E. Norris, N. P. Fiil, *EMBO J.* **3**, 909 (1984).
32. We thank R. Guillemin for encouragement, S. Reichlin for reading the manuscript and for useful suggestions, and G. Gravel and R. Schroeder for technical assistance. R.B. was supported by MRC grant MA 9145, by FRSQ and by The Montreal General Hospital Research Institute. N.L. and F.E. were supported by grants HD 09690 and AM I88II.

7 July 1987; accepted 25 September 1987

Synthesis of a Sequence-Specific DNA-Cleaving Peptide

JAMES P. SLUKA, SUZANNA J. HORVATH, MICHAEL F. BRUIST, MELVIN I. SIMON, PETER B. DERVAN

A synthetic 52-residue peptide based on the sequence-specific DNA-binding domain of Hin recombinase (139–190) has been equipped with ethylenediaminetetraacetic acid (EDTA) at the amino terminus. In the presence of Fe(II), this synthetic EDTA-peptide cleaves DNA at Hin recombination sites. The cleavage data reveal that the amino terminus of Hin(139–190) is bound in the minor groove of DNA near the symmetry axis of Hin recombination sites. This work demonstrates the construction of a hybrid peptide combining two functional domains: sequence-specific DNA binding and DNA cleavage.

MANY DNA-BINDING PROTEINS consist of two structural domains each with distinct functions. One domain recognizes and binds a specific DNA sequence, while the other catalyzes a chemical reaction on DNA. A synthetic 52-residue peptide identical to the COOH-terminal domain of Hin recombinase (190 amino acids) has been shown to bind to Hin recombination sites (26 bp) and to inhibit Hin activity (*1*). We now describe the conversion of this sequence-specific DNA-binding peptide into a sequence-specific DNA-cleaving peptide by covalent attachment of an iron chelator, ethylenediaminetetraacetic acid (EDTA), to the NH_2-terminus of the peptide. In the presence of Fe(II), molecular oxygen, and a reducing agent, the synthetic EDTA-peptide oxidatively cleaves DNA at Hin binding sites, revealing the base position and groove location of the NH_2-terminus of the peptide when bound to DNA.

X-ray crystal structures of a number of sequence specific DNA-binding proteins (or fragments) have become available (*2–7*). Comparison of the three-dimensional structures of λ-cro, λ-repressor, and catabolite gene activator protein (CAP) led to the postulate that a conserved α helix–turn–α helix motif is involved in sequence specific DNA recognition (*8, 9*). The crystal structure of a complex formed between the DNA-binding domain of phage 434 repressor (1–69) and a synthetic 434 operator DNA (14 bp) reveals that the peptide, which includes a helix-turn-helix domain, recognizes in DNA both a particular conformation and an array of base-pair contacts (*7*). Helix-turn-helix binding in the major groove of B-DNA may be a common structural feature for sequence-specific DNA affinity (*9, 10*).

Hin recombinase inverts a segment of DNA to change the expression of the flagellin genes of *Salmonella typhimurium* (*11*). Recombination occurs between two crossover sites, designated *hix*L and *hix*R in inverted repeat configuration, when they are on a supercoiled substrate (*12*). Each *hix* site is 26 bp long and has near twofold symmetry. Hin protein binds to a *hix* site as a dimer and protects bases −17 through +16 inclusive. Using solid-phase peptide synthesis, Bruist *et al.* (*1*) prepared a 52–amino acid peptide, identical to the COOH-terminus of Hin(139–190), that constitutes the putative sequence specific DNA-binding domain of Hin. The ability of Hin(139–190) to bind the *hix*L site was demonstrated by deoxyribonuclease (DNase) I and dimethyl sulfate protection experiments (footprinting) (*1*). The synthetic Hin(139–190) protected the same DNA sequence as Hin, except for the central three bases from −2 to +1. For purified Hin(139–190), an apparent dissociation constant of 0.2 μ*M* to *hix*L was determined by DNA mobility retardation assays (*13*).

Attachment of EDTA · Fe(II) to a DNA-binding molecule creates an efficient DNA-cleaving molecule (25°C, *p*H 7.0) (*14*). Attachment of EDTA converts a sequence-specific DNA-binding molecule to a bifunctional molecule capable of binding and cleaving DNA at the recognition site by oxidation of the DNA backbone (*15, 16*). Cleavage is initiated by the addition of a reducing agent such as dithiothreitol (DTT) or sodium ascorbate. Analysis of the cleavage patterns visualized by gel electrophoresis allows both the binding site location and the orientation of the synthetic DNA-binding molecules on double helical DNA to be identified (*15, 16*). Moreover, since the oxidative deoxyribose degradation is mediated by a diffusible species (most likely hydroxyl

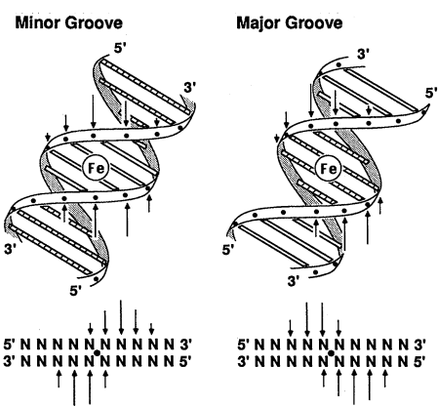

Fig. 1. Cleavage patterns produced by a diffusible oxidant generated by EDTA · Fe localized in the minor and major grooves of right-handed DNA. The edges of the bases are shown as open and crosshatched bars for the minor and major grooves, respectively.

J. P. Sluka and P. B. Dervan, Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125.
S. J. Horvath, M. F. Bruist, M. I. Simon, Division of Biology, California Institute of Technology, Pasadena, CA 91125.