Exons of the Human Pancreatic Polypeptide Gene Define Functional Domains of the Precursor*

Andrew B. Leiter‡, Mark R. Montminy§, Ellen Jamieson§, and Richard H. Goodman§
From the Divisions of §Gastroenterology and §Endocrinology, New England Medical Center Hospital, Boston, Massachusetts 02111

Pancreatic polypeptide is a 36-amino acid peptide which inhibits pancreatic exocrine function. We have previously determined from the nucleotide sequence of a cDNA that pancreatic polypeptide is derived from a 95-amino acid precursor, prepropancreatic polypeptide. Pulse-chase studies have suggested that the precursor is cleaved to produce three peptides: pancreatic polypeptide, an icosapeptide, and a smaller peptide. In the present study, we have used the cloned cDNA as a hybridization probe to isolate the pancreatic polypeptide gene from a human bacteriophage genomic library. The nucleotide sequence of 2.8 kilobases of DNA representing the entire human pancreatic polypeptide gene was determined. The gene contains four exons and three introns. Exon 1 encodes the 5'-untranslated region of the mRNA, exon 2 encodes the signal sequence and the sequence of pancreatic polypeptide, exon 3 encodes the icosapeptide, and exon 4 encodes a carboxyl-terminal heptapeptide and the 3'-untranslated region of the mRNA. By Southern blot analysis, the gene detected in a pancreatic polypeptide-producing islet cell tumor was indistinguishable from that in normal human leukocytes. The structure of the human pancreatic polypeptide gene is consistent with the hypothesis that prepropancreatic polypeptide generates three distinct peptides, each encoded by a separate exon. Increased expression of pancreatic polypeptide in the islet cell tumor does not appear to be correlated with major alterations in pancreatic polypeptide gene structure.

Pancreatic polypeptide is a 36-amino acid hormone secreted by cells within the endocrine and exocrine pancreas. Its secretion is stimulated by a variety of nutritive, metabolic, and neurogenic factors including fasting, protein meals, and hypoglycemia (1). Many of these stimuli appear to be mediated by cholinergic pathways. Pancreatic polypeptide inhibits secretion of enzymes and bicarbonate from the exocrine pancreas and causes relaxation of the gall bladder, but its effects on the endocrine pancreas are unclear (2). The limited effect of pancreatic polypeptide on insulin, somatostatin, and glucagon secretion, despite the proximity of PP cells to A, B, and D cells within the islets, has led some investigators to hypothesize that products of PP cells other than pancreatic polypeptide itself may have effects on islet hormone secretion. Like other peptide hormones, pancreatic polypeptide is synthesized as a larger precursor. Schwartz et al. (3, 4) have demonstrated by pulse-chase analysis that the prohormonal form of canine pancreatic polypeptide is post-translationally processed to produce pancreatic polypeptide, an icosapeptide, and a smaller third peptide derived from the carboxyl terminus of the precursor. The canine icosapeptide as well as its human counterpart have been isolated and sequenced (5), but to date no effect of this peptide on insulin, somatostatin, or glucagon secretion has been reported.

The nucleotide sequences of cDNAs representing the human pancreatic polypeptide precursor (6–8) have confirmed the studies of Schwartz et al. (3, 4). These sequences have indicated that pancreatic polypeptide is synthesized as a 95-amino acid precursor. Following the co-translational removal of a 29-amino acid signal sequence, a 66-amino acid prohormonal form is generated that contains the sequence of pancreatic polypeptide at its amino terminus (6). This molecule is further cleaved to produce the 36-amino acid pancreatic polypeptide. The fate of the carboxyl-terminal 27 amino acids of the precursor is uncertain. Possibly, this 27-amino acid peptide is cleaved at an arginine residue to produce the icosapeptide isolated by Schwartz et al. (5) and the additional carboxyl-terminal peptide characterized in the pulse-chase experiments. Alternatively, generation of the smaller peptide may occur during peptide isolation.

Genes encoding peptide hormones that undergo post-translational proteolytic cleavages are frequently divided into regions that correspond to functional domains in the precursor. Exon-intron junctions occur near the proteolytic cleavage sites in a number of hormone genes including gastrin (9), glucagon (10, 11), parathyroid hormone (12), cholecystokinin (13), and calcitonin (14). In order to better understand the biosynthesis of pancreatic polypeptide, we have characterized the structure of the gene and report its nucleotide sequence. The results of this study indicate that the intervening sequences within the pancreatic polypeptide gene occur at the putative cleavage sites in the precursor. These findings are consistent with the hypothesis that prepropancreatic polypeptide generates three distinct peptides.

EXPERIMENTAL PROCEDURES
Isolation of Recombinant Phage Containing the Human Pancreatic Polypeptide Gene—A human genomic library prepared from human liver DNA (15) was screened by the method of Benton and Davis (16) using a 350-base pair pancreatic polypeptide cDNA (6) as a hybridization probe. The cDNA was labeled to a specific activity of $10^9$ cpm/µg by nick translation (17). Two bacteriophage recombinants, XPP2A and XPP4A, contained DNA that hybridized to the probe. DNA was isolated from the recombinant phage by the method of Blattner et al. (18).

Restriction Endonuclease Analysis of the Purified Recombinant...
Bacteriophage DNA—Each phage DNA preparation was digested with several restriction enzymes, and the fragments were separated by agarose gel electrophoresis. Those fragments containing the pancreatic polypeptide gene were identified by Southern blot hybridization (19). A single 4.5-kb EcoRI restriction fragment of XPP4A and a 1.4-kb EcoRI fragment of APP4A hybridized to the probe. A hybridizing EcoRI fragment from each recombinant phage was isolated from an agarose gel and was ligated to EcoRI-digested pUC12 using T4 DNA ligase. The ligated DNA was transfected into Escherichia coli strain MC1061. DNA from one recombinant plasmid, pPP4A, was amplified in bacteria and isolated in sufficient quantity for nucleotide sequence analysis.

Nucleotide Sequence Analysis—Restriction fragments were labeled at either the 5' end with [γ-32P]ATP and T4 polynucleotide kinase or at the 3' end with the large fragment of DNA polymerase I and an α-32P-labeled deoxyribonucleotide and were sequenced by the method of Maxam and Gilbert (20).

Mapping of Transcriptional Initiation Site—RNA was prepared from a human pancreatic polypeptide-producing tumor by the method of Chibnall et al. (21) and was enriched for the polyadenylated fraction by affinity chromatography on oligo(dT)-cellulose. The 5' end of the mRNA was mapped with S1 nuclease. A hybridization probe was prepared by digesting pPP4A with the restriction enzyme HinfI and end labeling the fragments with T4 polynucleotide kinase. The labeled DNA was electrophoresed on a 5% acrylamide gel and a 100-base pair fragment was recovered from the gel by electroelution. The fragment was denatured by boiling and allowed to hybridize to the tumor RNA overnight at 57 °C in 80% formamide. The samples were then dried and digested with 65 units of S1 nuclease (Bethesda Research Laboratories) at 30 °C for 30 min (22). The sizes of the S1 nuclease-protected fragments were determined by electrophoresis alongside a sequencing ladder on an 8% denaturing acrylamide gel.

Southern Blot Analysis—Five to 10 μg of genomic DNA from human leukocytes or from the in vitro cell tumor were digested with various restriction enzymes and were electrophoresed on 0.7–1.0% agarose gels at 30–50 V. The fragments were transferred to nitrocellulose paper and allowed to prehybridize for 2 h in 1 × SET (1 × SET = 0.15 M NaCl, 0.05 M Tris-HCl, pH 8.0, 0.001 M EDTA) containing 50 μg/ml heparin and 0.5% SDS. The blot was then hybridized for 16 h at 60 °C in 1 × SET containing 0.5% SDS, 500 μg/ml heparin, and 10% (w/v) polyethylene glycol 8000 using a 1950-base pair XhoI-HindIII fragment extending from -300 to +1650 as a hybridization probe. Blots were washed at 60 °C in 1 × SET containing 0.5, 0.25, 0.125, and 0.038% SDS for 30 min each and in 0.2 × SET containing 0.1% SDS at 65 °C for 1 h. Following a final 30-min wash in 0.1 mm Tris base at 25 °C, the filter was exposed to x-ray film and the hybridizing bands were detected by autoradiography.

RESULTS

Isolation of a Bacteriophage Clone Encoding Human Pancreatic Polypeptide—Of 150,000 recombinant phage plaques screened, two contained DNA sequences that hybridized to the pancreatic polypeptide cDNA. The hybridizing phage were plaque-purified, and the phage DNA was digested with the restriction enzyme EcoRI. Southern blot hybridization of the digests from phage APP4A revealed a single 4.5-kb fragment that hybridized to the pancreatic polypeptide cDNA probe. This fragment was subcloned and mapped with a variety of restriction enzymes. The restriction map is depicted in Fig. 1.

Comparison of Cloned and Genomic DNA by Restriction Analysis—To verify that the cloned DNA was an authentic representation of the pancreatic polypeptide gene, we digested DNA from normal human leukocytes with the restriction enzymes PvuII, SphI, and NcoI and analyzed the fragments by Southern blotting using a cloned pancreatic polypeptide DNA probe. This fragment was subcloned and mapped with a variety of restriction enzymes. The restriction map is depicted in Fig. 1.

Nucleotide Sequence Analysis of the Cloned Pancreatic Polypeptide Gene—Approximately 2.8 kilobases of the 4.5-kb subcloned gene were sequenced. Comparison of the gene sequence to the sequence of the cDNAs revealed that the coding region is divided into four exons separated by three introns of 750, 245, and 199 bases (Figs. 1 and 3). Exon 1 encodes the 5'-untranslated region, extending from the transcriptional origin to the translational initiator methionine. The signal peptide and pancreatic polypeptide sequence are contained in exon 2. Exon 3 encodes the carboxy-terminal heptapeptide which flanks the sequence of pancreatic polypeptide. The carboxy-terminal heptapeptide as well as the 3'-untranslated region of the mRNA are contained in exon 4. The nucleotide sequence of the coding regions of the cloned gene is identical to the sequence of two cDNA hybridization probe (data not shown), suggesting that the additional hybridization arose from DNA sequences in one of the intervening sequences or the 5'-flanking region. We conclude that XPP4A contains a normal non-rearranged copy of the pancreatic polypeptide gene. In addition, Southern blots of DNA from the pancreatic polypeptide-producing tumor were indistinguishable from those of normal leukocyte DNA. Increased expression of pancreatic polypeptide in this tumor does not appear to be associated with major alterations in gene structure.

Nucleotide Sequence Analysis of the Cloned Pancreatic Polypeptide Gene—Approximately 2.8 kilobases of the 4.5-kb subcloned gene were sequenced. Comparison of the gene sequence to the sequence of the cDNAs revealed that the coding region is divided into four exons separated by three introns of 750, 245, and 199 bases (Figs. 1 and 3). Exon 1 encodes the 5'-untranslated region, extending from the transcriptional origin to the translational initiator methionine. The signal peptide and pancreatic polypeptide sequence are contained in exon 2. Exon 3 encodes the carboxy-terminal heptapeptide which flanks the sequence of pancreatic polypeptide. The carboxy-terminal heptapeptide as well as the 3'-untranslated region of the mRNA are contained in exon 4. The nucleotide sequence of the coding regions of the cloned gene is identical to the sequence of two cDNA hybridization probe (data not shown), suggesting that the additional hybridization arose from DNA sequences in one of the intervening sequences or the 5'-flanking region. We conclude that XPP4A contains a normal non-rearranged copy of the pancreatic polypeptide gene. In addition, Southern blots of DNA from the pancreatic polypeptide-producing tumor were indistinguishable from those of normal leukocyte DNA. Increased expression of pancreatic polypeptide in this tumor does not appear to be associated with major alterations in gene structure.

Nucleotide Sequence Analysis of the Cloned Pancreatic Polypeptide Gene—Approximately 2.8 kilobases of the 4.5-kb subcloned gene were sequenced. Comparison of the gene sequence to the sequence of the cDNAs revealed that the coding region is divided into four exons separated by three introns of 750, 245, and 199 bases (Figs. 1 and 3). Exon 1 encodes the 5'-untranslated region, extending from the transcriptional origin to the translational initiator methionine. The signal peptide and pancreatic polypeptide sequence are contained in exon 2. Exon 3 encodes the carboxy-terminal heptapeptide which flanks the sequence of pancreatic polypeptide. The carboxy-terminal heptapeptide as well as the 3'-untranslated region of the mRNA are contained in exon 4. The nucleotide sequence of the coding regions of the cloned gene is identical to the sequence of two cDNA hybridization probe (data not shown), suggesting that the additional hybridization arose from DNA sequences in one of the intervening sequences or the 5'-flanking region. We conclude that XPP4A contains a normal non-rearranged copy of the pancreatic polypeptide gene. In addition, Southern blots of DNA from the pancreatic polypeptide-producing tumor were indistinguishable from those of normal leukocyte DNA. Increased expression of pancreatic polypeptide in this tumor does not appear to be associated with major alterations in gene structure.
of the cDNAs previously reported (6, 7). The single base substitution in the coding sequence reported by Takeuchi and Yamada (8) was not present in the cloned gene and may represent a polymorphism. Characteristic G-T/A-G donor acceptor sites were present at the ends of all three intervening sequences. A region of alternating purine and pyrimidine nucleotides was found between positions +242 to +296 within the first intron. Similar sequences have been found to adopt a Z-DNA structure (23).

**Promoter, Transcriptional Initiation Site, and 5'-Flanking Sequences of the Human Pancreatic Polypeptide Gene**

The transcriptional initiation site was determined by S1 nuclease mapping of mRNA from a human islet cell tumor (Fig. 4). Two DNA fragments differing in length by a single nucleotide were resistant to S1 nuclease digestion. These studies localized the transcriptional initiation point to a site 60-61 base pairs upstream from the translational start codon. The degeneracy of the transcriptional origin deduced by S1 mapping could arise from one of several causes including degeneracy of the cap site, mRNA degradation, or instability of mRNA-DNA duplexes close to the cap. The transcriptional initiation site was 24-25 bases downstream from a canonical sequence, TATAAA. The nucleotide sequence of an additional 900 bases upstream from the promoter was also determined and is shown in Fig. 3. Potential consensus core enhancer-like sequences and possible regions of Z-DNA structure are underlined.

**Repetitive DNA Sequences**

Repetitive DNA sequences were initially identified by Southern blot hybridizations of subcloned DNA digested with the restriction enzyme MstII. The blots were probed with nick-translated human genomic DNA.
DNA. These studies revealed hybridization to a 1.6-kb fragment which extended from intron 3 to 1.4 kb beyond the polyadenylation site. The repetitive element was further localized by blot hybridizations of subcloned DNA digested with SphI and EcoRI or HindIII and BamHI. This analysis localized the repetitive element to a sequence between 700 and 1500 bases downstream from the polyadenylation site (see Fig. 1).

**DISCUSSION**

In the present study, we report the nucleotide sequence of the human prepropancreatic polypeptide gene isolated from a bacteriophage genomic library. Restriction endonuclease analysis of a total human genomic DNA indicates that the cloned DNA fragment is an authentic representation of the gene. Characteristic features of the pancreatic polypeptide gene include a Goldberg-Hogness promoter consensus sequence separated by 24 or 25 bases from the cap site, three intervening sequences, and a repetitive DNA element between 700 and 1500 bases downstream from the polyadenylation site (see Fig. 1).

The coding regions of most hormone genes are interrupted by intervening sequences, as shown schematically in Fig. 5 (9-14, 27, 31). In some instances, such as in the somatostatin (32) or pro-opiomelanocortin (33) genes, the intervening sequences do not appear to divide the hormone precursor into functional domains. The prepropancreatic polypeptide gene, on the other hand, is clearly divided into functional domains by its intervening sequences. Like many of the other hormone genes depicted in Fig. 5, the pancreatic polypeptide gene contains an intervening sequence near the translational initiation site. In the case of pancreatic polypeptide, this intron occurs precisely at the initiator AUG codon, thereby separating the ribosomal binding site from the protein coding region. Furthermore, the structure of the prepropancreatic polypeptide gene indicates that the signal region of the precursor is not specifically demarcated by intervening sequences. As shown in Fig. 5, the signal regions of most hormone precursors also do not appear to be encoded by discrete structural units of the respective genes.

The division of hormone coding regions into structural and/or functional domains by intervening sequences may serve several functions in generating the diversity of peptide hormones. Separating the coding regions for discrete structural domains within a polypeptide may permit conservation of one domain while allowing divergence of another (34). Such a process may have occurred during the evolution of preprotocagon (10, 11), for example. Separation of functional regions in individual exons may also provide a mechanism to increase diversity through alternative RNA splicing. This mechanism appears to be used by the calcitonin (14) and tachykinin (35) genes.
FIG. 5. Structural organization of selected peptide hormone genes. Structures of selected peptide hormone precursors are shown schematically. Protein coding regions are shown as boxes. The positions of intervening sequences in the genes are indicated by inverted solid triangles. Putative post-translational cleavage sites are shown as arrows below each figure. SP, signal peptide; 5′, 5′-untranslated region of mRNA; 3′, 3′-untranslated region of mRNA; PP, pancreatic polypeptide; aa, amino acids; GLUC, glucagon; GLP, glucagon-like peptide; VIP, vasoactive intestinal polypeptide; PHM, peptide with N-terminal histidine and C-terminal methionine; CCK, cholecystokinin; PTH, parathyroid hormone; A, B, and C, the A, B, and C peptides of rat preproinsulin II; G17, gastrin 17; CAL, calcitonin; CCP, calcitonin cleavage product; CGRP, calcitonin gene-related peptide.

Because the biological functions of the icosenopeptide and heptapeptide derived from the pancreatic polypeptide precursor are unknown, it is unclear whether the introns divide the gene into functional, in addition to structural, domains. As mentioned above, the interruption of hormone genes by introns may allow independent accumulation of changes within portions of a precursor. Therefore, in the case of propancreatic polypeptide, it might be expected that some portions of the precursor would be more highly conserved than others. The sequences of human, ovine, and canine pancreatic polypeptide differ in only four positions out of 36. The human, ovine, and canine icosenapeptides vary at 12 of the 20 positions (5). The poor conservation of the icosenapeptide suggests that this peptide may not have an essential biological function or that its function may differ among species. No data are available at present regarding conservation of the heptapeptides at the carboxyl terminus of the precursor. A complete understanding of the physiological significance of the pancreatic polypeptide gene will require the elucidation of the functional role of both the icosenapeptide and heptapeptide.

Acknowledgments—We thank T. Krontiris for help with the Southern blots, T. Tsukada and G. Mandel for valuable discussions, and K. Kelley for preparation of this manuscript.

REFERENCES

1. Hazelwood, R. L. (1981) in The Islets of Langerhans (Cooperstein, S. J., and Watkins, D., eds) pp. 273–317, Academic Press, New York

2. Adrian, T. E., Greenberg, G. R., and Bloom, S. R. (1981) in Gut Hormones (Bloom, S. R., and Polak, J. M., eds) 2nd Ed., pp. 206–212, Churchill-Livingston, Inc., New York

3. Schwartz, T. W., Gingerich, R. L., and Tager, H. S. (1980) J. Biol. Chem. 255, 11494–11498

4. Schwartz, T. W., and Tager, H. S. (1981) Nature 294, 589–591

5. Schwartz, T. W., Hansen, H. F., Hakanson, R., Sundler, F., and Tager, H. S. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 708–712

6. Leiter, A. B., Keutmann, H. T., and Goodman, R. F. (1984) J. Biol. Chem. 259, 14702–14706

7. Boel, E., Schwartz, T. W., Norris, K. E., and Filil, N. P. (1984) EMBO J. 3, 909–912

8. Takeuchi, T., and Yamada, T. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 1536–1539

9. Wiborg, O., Berglund, L., Boel, E., Norris, R., Norris, K., Rehfeld, J. F., Marcker, K. A., and Vuuat, J. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 1067–1069

10. Bell, G. I., Sanchez-Pescador, R., Laybourn, P. J., and Najarian, R. C. (1983) Nature 304, 368–371

11. Heinzlrich, G., Gros, P., and Habener, J. F. (1984) J. Biol. Chem. 259, 14082–14087

12. Vonderhaar, T. J., McDevitt, B. E., Freeman, M. W., Fennick, B. J., Hendy, G. N., Potts, J. T., Rich, A., and Kronenberg, H. M. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 2127–2131

13. Deschenes, R. J., Haun, R. S., Funckes, C. L., and Dixon, J. E. (1985) J. Biol. Chem. 260, 1280–1286

14. Rosenfeld, M. G., Amara, S. G., and Evans, R. M. (1984) Science 225, 1315–1319

15. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G., and Maniatis, T. (1978) Cell 15, 1157–1174

16. Benton, W. D., and Davis, R. (1977) Science 196, 180–182

17. Rigley, R. W., Dieckmann, M., Rhodes, C., and Berg, P. (1977) J. Biol. Chem. 251, 237–251

18. Blattner, F. R., Williams, B. G., Blechl, A. E., Deniston-Thompson, K., Fater, H. E., Furlong, L. A., Gurnwald, D. S., Keifer, D., Moore, D., Sheldon, E. L., and Smithies, O. (1977) Science 196, 161–163

19. Southern, E. M. (1975) J. Mol. Biol. 98, 503–517

20. Maxam, A., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 560–564

21. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) Biochemistry 18, 5294–5299

22. Weaver, R. F., and Weissman, C. (1979) Nucleic Acids Res. 7, 1175–1185

23. Nordeinheim, B., and Rich, A. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 1821–1825

24. Goodman, R. H., Jacobs, J. W., Dee, P. C., and Habener, J. F. (1982) J. Biol. Chem. 257, 1156–1159

25. Weiler, H., Konig, M., and Gruss, P. (1983) Science 219, 626–631

26. Schwartz, T. W. (1979) Scand. J. Gastroenterol. 14, Suppl. 53, 93–100

27. Tsukada, T., Horovitch, S. J., Montminy, M. R., Mandel, G., and Goodman, R. H. (1985) DNA (N. Y.) 4, 293–300

28. Tatemoto, K. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 5485–5489

29. Tatemoto, K. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 2514–2518

30. Minth, C. D., Bloom, S. R., Polak, J., and Dixon, J. E. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 4577–4581

31. Lomedico, P., Rosenthal, N., Eistratiadis, A., Gilbert, W., Kolodner, R., and Tizard, R. (1978) Cell 18, 545–558

32. Montminy, M. R., Goodman, R. H., Horovitch, S. J., and Habener, J. F. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 3337–3340

33. Takahashi, H., Hakamata, Y., Watanabe, Y., Kikuno, R., Miyata, T., and Numa, S. (1983) Nature 302, 6847–6858

34. Gilbert, W. (1978) Nature 271, 501

35. Nawa, H., Kotani, H., and Nakaniishi, S. (1984) Nature 312, 729–734